TECHNICAL MEMORANDUM

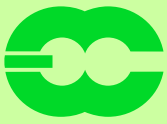
679

ECMWF study to quantify the interaction between terrestrial and space-based observing systems on Numerical Weather Prediction skill

Gábor Radnóti, Peter Bauer, Anthony McNally and András Horányi

Research Department

July 18, 2012



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

The subject of the present study is the evaluation of the interaction between terrestrial and space-based observing systems and thus focuses on the inter-dependence that is common to all current operational NWP systems.The study also addresses selected recommendations that were issued by the WMO expert team on the evolution of the global observing system (ET-EGOS) at its meeting in December 2009. To reach this goal different Observation System Experiments (OSEs) have been perfomed and evaluated. The main work packages presented in the study are:

- Investigation of the impact of a thinned terrestrial observing system on radiance bias correction anchoring (OSEs with and without GPSRO and with full and less dense conventional observing system) with special focus on:
  - the synergy between radiosondes and GPSRO as anchors (with resepect to temperature);
  - the question of what radiosonde coverage is needed in the stratosphere, to which height, and for which latitude ranges;
  - the synergy between ships/buoys and GPSRO as anchors (with respect to surface pressure).
- Investigation of the impact of a reduced conventional observing system on NWP following the most successful scenarios (3b) and (4) of an earlier Upper Air Network Redesign Study (Radnoti, 2010). In this scenario (3b) 06, 12 and 18 UTC radiosonde ascents are reduced by removing soundings of sites which are in the vicinity of in a 100 km radius in the vicinity of airports in Europe. In scenario (4) the horizontal spacing of sites is 250 km, all other settings are similar to scenario (3b).
- Investigation of the impact of ASAP radiosondes over the Northern Atlantic region.

Contents

1 Executive summary

The subject of the present study is the evaluation of the interaction between terrestrial and space-based observing systems and thus focuses on the inter-dependence that is common to all current operational NWP systems. The expected results are to better understand and quantify this inter-dependence with a special focus on selected scenarios and observing system configurations. The study also addresses selected recommendations that were issued by the WMO expert team on the evolution of the global observing system (ET-EGOS) at its meeting in December 2009.

In the current ECMWF four-dimensional assimilation system the model in the troposphere is assumed to be perfect over the 6-hour/12-hour assimilation window so that model errors and model biases are not explicitely accounted for. In the stratosphere, model error is accounted for but not cycled through successive analyses and thus explicitly accounting for it is of little impact.

Systematic differences between model and observations are treated by quality control and bias correction. The quality control excludes data in situations in which model and/or observation operators are considered too inaccurate. The bias correction (Dee 2005) eliminates systematic differences between model and observations for the data, that passed the quality control, in a dynamic and state-dependent way. The assumption of a perfect model implies that the bias correction is only applied to the observations.

Only selected conventional observations are corrected, namely surface pressure observations (Vasiljevic et al. 2005) and the solar radiative heating contribution to radiosonde temperature day-time biases (Vasiljevic et al. 2007). In the future, AMDAR observations will be corrected as well since substantial differences between ascending and descending temperature profiles over the same location have been found.

A variational bias correction is applied to all radiance observations (about 95% of the assimilated data), total column ozone and total column water vapour observations. The dynamic nature of this correction requires complementary observations that provide anchoring points so that the bias correction does not excessively absorb model biases (e.g. Auligné et al. 2007). Currently, these anchoring points are represented by conventional and GPS radio occultation (GPSRO) observations. Therefore, the optimal utilization of the bulk of the observational data relies on a steady and well defined set of this data. This issue has been identified by the WMO ET-EGOS as one of the priority topics to be studied by observing system experiments (OSEs).

Note that systems like this are currently run at most operational NWP centres so that the topics addressed in this study represent issues related to a wider community.

The main work packages covered in this study aim:

- To investigate the impact of a thinned terrestrial observing system on radiance bias correction anchoring (OSEs with and without GPSRO and with full and less dense conventional observing system) with special focus on:
  - the synergy between radiosondes and GPSRO as anchors (with resepect to temperature);
  - the question of what radiosonde coverage is needed in the stratosphere, to which height, and for which latitude ranges;
  - the synergy between ships/buoys and GPSRO as anchors (with respect to surface pressure).
- To investigate the impact of a reduced conventional observing system on NWP following the most successful scenarios (3b) and (4) of the Upper Air Network Redesign Study (Radnoti, 2010). In

this scenario (3b) 06, 12 and 18 UTC radiosonde ascents are reduced by removing soundings of sites which are in the vicinity of in a 100 km radius in the vicinity of airports in Europe. In scenario (4) the horizontal spacing of sites is 250 km, all other settings are similar to scenario (3b).

- To investigate the impact of ASAP radiosondes over the Northern Atlantic region.

Aircraft and radiosonde impact studies with different GPSRO data coverages

The first set of OSEs has been performed to investigate the interaction of radiosonde, aircraft and GPSRO measurements in terms of their analysis and forecast influence. The synergy of these observations as anchors for the bias correction of other observations was also studied. A newly developed tool to verify forecasts against conventional and satellite observations was applied and it proved to give important additional information on the observation impacts. The set of OSEs consisted of radiosnde and aircraft denial experiments performed with successively reduced GPSRO density. Stratospheric radiosonde impacts were additionally studied with denial of radiosonde data above 50hPa.

The aircraft denial experiments showed a remarkable influence on the mean analysis state. As expected, this is the strongest near the cruise level (around 250hPa) over the Northern hemisphere, where the change of mean temperature analysis reaches 0.4K. This is clearly related to a warm temperature bias of aircraft measurements that can be also detected in the analysis and first fuess fit to GPSRO bending angles. In spite of the strong temperature biases, aircraft data prove to have a significantly positive impact on forecast scores.

Radiosonde data also have an imapct on the mean analysis state, mostly over large continental areas. The temperature impact is a warming in the boundary layer and a cooling effect above. Near the tropopause the impact is small in spite of the fact that mean radiosonde temperature departures are the largest there. This is likely due to the overwhelming dominance of aircraft data near the cruise level. Radiosonde assimilation has a clearly visible impact on the mean humidity state too, that is an overall moistening of the analysis over most of the continents. This is the integral result of a dominant moistening in the lower troposphere and a drying impact above. Radiosonde data, just like aircraft measurements, have a significant positive impact on forecast scores. Here the positive impact is observed for humidity scores as well.

The comparison of aircraft and radiosonde impact on forecast scores suggests that in the higher troposphere aircraft data contribute more to the forecast score improvement and this dominance spreads down to 400-500hPa. Below this, radiosondes are more beneficial on a hemispheric average. Nevertheless, over Europe and North-America, where the largest observation density of both radiosonde and aircraft data profiles are available, the aircraft contribution to the forecast quality is dominant even in the lower troposphere. Stratospheric radiosonde data prove to have a particularily strong positive impact on wind forecast scores, mainly over the Tropics. This was proven both against analysis and against radiosonde data.

As it has been shown in earlier studies, GPSRO data also strongly constrain temperature profiles and surface pressure of the analysis. Therefore it is interesting to see the impact of radiosonde and aircraft data in conjunction with GPSRO data coverage. The OSEs with different GPSRO data densities showed that in the stratosphere GPSRO data have a warming effect while in the troposphere they cool the analysis. This impact is in agreement with radiosonde temperature profiles except for the lower boundary layer. Comparison of GPSRO impacts in conjunction with aircraft and radiosonde denials suggests the impact of GPSRO data on the mean analysis state is mainly driven by aircraft biases.

The synergy between radiosondes and GPSRO as temperature anchors has been examined through the denial experiments. It has been found that the presence or absence of radiosondes did not show strong impact on the bias correction, at least much less than GPSRO data. The small, but still detectable anchoring effect of radiosondes proved to be stronger when less GPSRO data were assimilated.

Follow-on investigations of the Upper air redesign study

The EUCOS Upper-air redesign study (Radnoti 2010) has concluded that two data density scenarios that deserve the most attention are Sc3b and Sc4, i.e. a combined radiosonde/aircraft profile density of 100 km and 250 km, respectively. The same evaluation of these scenarios has been performed for a 6 month continuos period. For most of the parameters and vertical levels scenario-3b did not show significant difference with respect to the control experiment in agreement with the findings of the 2009 study. The only parameter where both verification against operational and against own analysis showed a significant deterioration of the forecast due to data denial is 100hPa wind in the first 48h. For scenario-4 the impact is significant for wind and temperature forecasts throughout the Troposphere and for low level humidity. The significant impact lasts 2-3 days. The impact of radiosonde and aircraft measurement density on the performance of precipitation forecast has also been quantified for the entire 6 months of experimentation. Precipitation forecasts were compared to rain gauge measurements over Western Europe. Overall, the data denial has little effect on the precipitation scores and it is within the sampling uncertainty. The only systematic effect is seen for the 12UTC run at day 1.

Marine buoy and VOS surface observation impact studies

The impact of the marine buoy and VOS surface observations has been examined by running another set of OSEs. A special emphasis has been put on the impact of North-Atlantic marine buoys. A two-month OSE from December 2008 to January 2009 has been performed to assess the impact of the extra surface pressure measurements that were introduced during the last decade in the framework of the EUCOS E-Surfmar program. The control experiment used the operational observing system. The OSE has been run with the standard 4D-Var system using T511 horizontal resolution and 91 vertical levels. The impact of the global denial of buoy data has been also investigated. Both the control and the denial experiments have been repeated without GPSRO radio occultation data, since GPSRO assimilation has been recently proven to constrain the surface pressure analysis. Later the experiment pair of partial denial without GPSRO data has been repeated for a 6 month period to see the significance of the relatively small impact of partial data denial. The forecast impact of buoy and GPSRO data has been examined in two severe winter storm cases (storms Klaus and Xynthia), that recently affected Europe. The following conclusions can be drawn form the performed experiments:

- The impact of buoy surface pressure observations on the forecast performance is large, especially in the lower atmosphere and it is more expressed over the Southern Hemisphere.
- This impact is more evident when GPSRO data are not assimilated: significant impact lasts 2-3 days with and 4 days or more without assimilation of GPSRO data.
- The comparison of the surface pressure forecast score impact of buoy and GPSRO data shows that buoy data are more important until 24h while from 48h onwards GPSRO data clearly contirbute more to the forecast quality. when going higher in the atmosphere GPSRO data become more and more dominating over buoy data.

- Additional buoy data introduced at the North-Atlantic area within the E-SURFMAR program prove to locally improve surface pressure forecast scores (i.e. for the Northern Atlantic domain, but also for Europe) and it lasts up to 1-4 days.
- The impact of the additional buoy data can be clearly seen in extreme weather events.
- To detect impact of the additional data on precipation scores is diffiult, however some moderate positive impact can be identified.

ASAP impact studies

The impacts of the ASAP shipborne radiosondes were also investigated for a period of 3 months (July-September, 2011) in order to assess the importance of the ASAP radiosondes in the forecasts of the IFS global model. The applied model resolution was T511 with the use of 91 model levels. The verification data was used from the operational IFS analysis, the analysis of the experiments and the observations, respectively. The main conclusions of the investigations are birefly summarised as follows:

- Slight, but mostly not significant improvements can be shown at the first 2 days forecasting range for the Northen Atlantic area (especially if the verification is realised against the IFS operational analysis).
- At the 3-4 days forecasting range rather slight degradation is seen, the missing data has a positive contribution to the IFS forecasts.
- The precipitation scores are slightly, but clearly negative especially for the 12 UTC integration at the 3-5 days forecast ranges.
- The obtained results are therefore rather ambiguous, i.e. no clear positive impact of the additional ASAP radiosondes can be demonstrated, moreover more scores indicate negative impact. It is advised to extend the experimentation for a longer period for getting more robust conclusions.

Acronyms

AMDAR	Aircraft Meteorological Data Relay
AMSR-E	Advanced Scanning Microwave Radiometer - E
AMSU	Advanced Microwave Sounding Unit
ASAP	Automated Shipboard Aerological Programme
DCDA	Delayed cut-off Daily Analysis
ECMWF	European Centre for Medium-Range Weather Forecasts
EPS	EUMETSAT Polar System
ETS	Equitable Threat Score
E-SURFMAR	Surface Marine observation programme of EUMETNET
EUCOS	EUMETNET Composite Observing System
EUMETNET	Network of European Meteorological Services
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
ET-EGOS	Expert Team on Evolution of Global Observing System
FEC	Forecast Sensitivity to Observations
GCOS	Global Climate Observing System
GPS	Global Positioning System
GPSRO	GPS Radio Occultation
GUAN	GCOS Upper-air Network
HIRS	High-Resolution Infrared Radiation Sounder
IASI	Infrared Atmospheric Sounding Interferometer
IFS	Integrated Forecasting System
MHS	Microwave Humidity Sounder
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OSE	Observing System Experiment
PSS	Peirce Skill Score
RMSE	Root-Mean-Squarre Error
SEEPS	Stable Equitable Error in Probability Space
TCWV	Total Column Water Vapor
UCAR	University Corporation for Atmospheric Research
WMO	World Meteorological Organization
UTC	Coordinated Universal Time
VOS	Voluntary Observing Ship

2 Background

The skill of Numerical Weather Prediction (NWP) depends to a large extent on the quality of the analysis state that serves to initialize the forecast model. At ECMWF and most operational centres this is performed via a four-dimensional variational assimilation (4D-Var) scheme that produces a physically consistent estimate of the state of surface and atmosphere. This estimate is constrained by both the forecast model and observations. The large improvement of NWP skill obtained over the past two decades can be attributed to model developments and the increased capability of the observation system along with sophisticated data assimilation systems that derive the analysis state from model forecasts and data.

In October 2001, the Members of EUMETNET that are represented by the EUMETNET Council decided to join their efforts to optimise the ground-based meteorological observing system at European scale under the "EUMETNET Composite Observing System (EUCOS) Operational Programme". The EUCOS Operational Programme defined recommendations for the evolution of the EUCOS in the period 2002-2006 noting that an on-going Studies Programme would further refine and underpin the programme proposal. In 2006, the EUMETNET Council approved the programme proposal developed by DWD for the 2007-2011 period, which again contains plans for the EUCOS Studies Programme and specific plans for Observation System Experiments (OSE). A document describing the "EUCOS Upper-air Network Design Observing System Experiment 2007" has been drafted by EUCOS in 2007 and was approved by PB-OBS and EUMETNET Council.

The Space-Terrestrial studies, conducted between 2005 and 2007, have demonstrated the crucial contribution of the terrestrial observing system to NWP skill, in particular in the Northern hemisphere and over Europe. Regarding the space component a series of OSEs have been performed on behalf of EUMETSAT (Bauer and Radnoti 2009) in 2008-10, focusing on an impact estimation of a potential loss of EPS Metop instruments. The results emphasized the crucial contribution of Metop to the Global Observing System (GOS). In 2010, another study was performed (Radnoti et al. 2010) aiming at a more detailed assessment of different groups of spaceborne observing systems, related to temperature sounding, moisture/cloud/precipitation imaging, radio-occultation, wind and soil moisture observations. The results from this study indicated the sensitivity of quality control and bias correction of the space-based system to the availability of the terrestrial system suggesting a strong inter-dependence of the two observation networks.

The subject of the present study is the evaluation of the interaction between terrestrial and space-based observing systems and thus focuses on the above mentioned inter-dependence that is common to all current operational NWP systems. The expected results are to better understand and quantify this inter-dependence with a special focus on selected scenarios and observing system configurations as outlined in the next section. The study also addresses selected recommendations that were issued by the WMO expert team on the evolution of the global observing system (ET-EGOS) at its meeting in December 2009.

3 Study objectives and approach

The present study mainly focuses on the understanding and demonstration of the interaction between the impact of the terrestrial and the space-based observing systems. This interaction is most strongly related to how much different configurations of the observing system affect the mean state of the atmosphere and therefore the mean differences between observations and short-range model forecasts. These mean differences, or biases, can be substantial and require careful consideration in data assimilation to identify

contributions from the NWP model and observations and to avoid aliasing effects that can produce sub-optimal analyses that directly affect forecast skill.

In the current ECMWF four-dimensional assimilation system the model in the troposphere is assumed to be perfect over the 6-hour/12-hour assimilation window so that model errors and model biases are not explicitely accounted for. In the stratosphere, model error is accounted for but not cycled through successive analyses and thus explicitly accounting for it is of little impact.

Systematic differences between model and observations are treated by quality control and bias correction. The quality control excludes data in situations in which model and/or observation operators are considered too inaccurate. The bias correction (Dee 2005) eliminates systematic differences between model and observations for the data, that passed the quality control, in a dynamic and state-dependent way. The assumption of a perfect model implies that the bias correction is only applied to the observations.

Only selected conventional observations are corrected, namely surface pressure observations (Vasiljevic et al. 2005) and the solar radiative heating contribution to radiosonde temperature day-time biases (Vasiljevic et al. 2007). In the future, AMDAR observations will be corrected as well since substantial differences between ascending and descending temperature profiles over the same location have been found.

A variational bias correction is applied to all radiance observations (about 95% of the assimilated data), total column ozone and total column water vapour observations. The dynamic nature of this correction requires complementary observations that provide anchoring points so that the bias correction does not excessively absorb model biases (e.g. Auligné et al. 2007). Currently, these anchoring points are represented by conventional and GPS radio occultation (GPSRO) observations. Therefore, the optimal utilization of the bulk of the observational data relies on a steady and well defined set of this data. This issue has been identified by the WMO ET-EGOS as one of the priority topics to be studied by OSEs.

Note that systems like this are currently run at most operational NWP centres so that the topics addressed in this study represent issues related to a wider community.

The main work packages covered in this study aim:

- To investigate the impact of a thinned terrestrial observing system on radiance bias correction anchoring (OSEs with and without GPSRO and with full and less dense conventional observing system) with special focus on:
  - the synergy between radiosondes and GPSRO as anchors (with resepect to temperature);
  - the question of what radiosonde coverage is needed in the stratosphere, to which height, and for which latitude ranges;
  - the synergy between ships/buoys and GPSRO as anchors (with respect to surface pressure).
- To investigate the impact of a reduced conventional observing system on NWP following the most successful scenarios (3b) and (4) of the Upper Air Network Redesign Study (Radnoti, 2010). In this scenario (3b) 06, 12 and 18 UTC radiosonde ascents are reduced by removing soundings of sites which are in the vicinity of in a 100 km radius in the vicinity of airports in Europe. In scenario (4) the horizontal spacing of sites is 250 km, all other settings are similar to scenario (3b).
- To investigate the impact of ASAP radiosondes over the Northern Atlantic region.

The study will produce average performance statistics over a representative period and will also include an application of the observing system scenarios to selected extreme weather case studies. The

performance statistics will include the recently developed advanced diagnostics to complement model-observation statistics and standard forecast scores.

4 Configuration

4.1 Model set-up

The experiments have been run with model cycle CY36R4 of the Integrated Forecasting System (IFS) that became operational at ECMWF on 2010/11/09. This choice ensured that the latest upgrades to model physics, data assimilation and observation treatment were incorporated in the OSEs and the most recently developed diagnostic tool of forecast departure statistics could be applied in the evaluation of the results. The experiments have been run at reduced horizontal resolution, namely for the forecast model with a T511 wavenumber truncation (40 km, compared to operational resolution T1279 i.e. 16 km). Based on the findings from earlier OSEs the vertical resolution of the experiments has been kept at 91 levels with a model top level pressure at 0.01 hPa, as in operations. From previous OSEs, the reduced horizontal resolution was shown to produce sufficiently accurate results.

The system has been run in the global 4D-Var configuration that produces two analyses per day (00 and 12 UTC) with 12-hour assimilation windows (delayed cut-off data assimilation configuration, DCDA; Haseler 2004). Two medium-range forecasts have been run per day, again as in the operational system configuration. The main experimentation period is July-September 2008 because of the additional experimental GPSRO data available for this period (Radnoti et al. 2010).

All experiments have been initialized with the operational suite on the first day of the respective period. An initial 14-day spin-up phase was included allowing for the system to adjust to the modified observing system that was activated on day-1 in each case. It is expected that by the end of the spin-up period, the system has lost memory of the full system used to initialize the analysis on day-1. The experiment evaluation has been restricted to the remaining part of the period. The evaluation has been performed based on standard observation consistency statistics and standard forecast skill scores. The observation consistency statistics have been computed for two- and four-day forecasts in addition to the standard first-guess and analysis observation-minus-model statistics.

The variational bias-correction (Dee 2005, Auligné et al. 2007) for the experimentation period was initialized with the operational system output on the initial date and left active throughout the experimentation periods. This ensured that, as in the operational system, a trade-off between analysis and bias increments as a function of model state is performed.

4.2 Definition of impact evaluation

The evaluation of analyses is usually performed with all used observations assuming that better analyses will produce a generally and consistently better fit of the model fields when compared to observations. This comparison is performed for both analysis and first-guess, i.e. the short-range forecast that produces a first estimate of the actual state and that has been initialized from the previous analysis. The comparison is performed against all conventional and satellite observations and therefore uses the same observation operators (interpolation scheme, radiative transfer model) as the data assimilation scheme. This method is very stable and generally considered unambiguous.

A new development that is already available in the applied model version made it possible to extend

this method from first-guess and analysis to a forecast of any time range (Radnoti 2011). However, the derivation of these additional statistics requires a repeated screening run of the model for each time range of each individual forecast. 48 and 96 hour forecasts have been chosen for this extension in this study. This extension is especially important because, in the forecast evaluation, different reference standards can produce rather different results. Bauer and Radnoti (2009) have shown that the evaluation with the experiment's own analysis or the operational analysis could produce inconsistent results, in some cases with opposite signs. The argument for using the experiment's own analyses for evaluation is that if an additional observation is expected to change the mean analysis state, only the own analysis provides a fair reference while any other would represent a poorer analysis state. This, however, can also produce problems in case of a sub-optimal observing system and simply due to the fact that additional observations can add systematically larger increments to the analysis and therefore increase the root-mean-square difference between (short-range) forecasts and analyses. Standard forecast evaluation in this study is performed with the operational analyses and with observations, the latter applying the new forecast departure configuration.

Forecast scores or analysis statistics are often evaluated in this study as difference curves or difference fields. The standard for all these difference plots is Experiment – Reference, unless stated otherwise.

4.3 Observing system experiments

4.3.1 Global radiosonde and aircraft impact with different GPSRO constellations

The aim of this work package is to examine the global impact of radiosonde data as compared to aircraft observations with the main emphasis on the impact on the mean state. As earlier studies have proven, GPSRO observations give a very valuable contribution to the NWP observing system due to their absolute calibration (i.e. they are considered bias-free) and their strong sensitivity to upper atmospheric temperature structures. Because of this they do not need to be bias corrected and they may serve as anchors for temperature analyses. Therefore the radiosonde and aircraft data impacts are studied with different background GPSRO constellations to gradually change the GPSRO constraint.

An additional question is how much *stratospheric* radiosonde measurements at different latitudinal bands contribute to the quality of the forecasts both in the presence and absence of a dense GPSRO obseravtion sampling. This question is also addressed in this work package.

4.3.2 Follow-on investigations of the Upper air redesign study

The Upper-air redesign study (e.g. Radnoti 2010) of EUCOS has concluded that two data density scenarios that deserve the most attention are Sc3b and Sc4, i.e. a combined radiosonde/aircraft profile density of 100 km and 250 km, respectively. In this work package the data denial experiments corresponding to these two scenarios are revisited with the new model version and for the continuous 3-month period of this study.

4.3.3 Buoy and ship measurement impact studies

In this work package the impact of buoy and ship measurements in the presence and absence of dense GPSRO data coverage is examined. The synergy between ships/buoys and GPSRO as anchors is studied by data denial experiments and different ways of surface pressure bias treatment for buoy measurements.

4.3.4 ASAP impact study

In this study the impact of the ASAP radiosondes is explored, i.e. it is quantified how the ASAP soundings over the Northern Atlantic influence the forecasts of the IFS model. It is noted here that although the total amount of ASAP data is not very high, but they are located at a region, where the observation density is rather low especially as far as the vertical profiles are concerned.

5 Results

5.1 Radiosonde, aircraft and GPSRO impact studies

5.1.1 Global radiosonde and aircraft impact in different GPSRO data coverages

The earlier Space-Terrestrial impact studies (e.g. Thepaut and Kelly 2007) have shown that additional radiosondes (in addition to those in the so called GUAN radiosonde network) and aircrafts (temperature and wind) contribute more or less equally to the observing system (with a slight advantage for radiosonde) and that these two observing systems are complementary.

In the present work package this issue is revisited, but in a slightly different way: first, impact of both observing systems is examined against a total data withdrawal of the given observation type. Secondly, the main emphasis is put on the impact on mean analysis state and change of bias properties as seen from the perspective of the variational bias correction applied in the IFS model to most satellite observations.

Further, radiosonde and aircraft data impact is compared to GPSRO impact and the dependence of the radiosonde and aircraft data impact on GPSRO data coverage is examined. As mentioned before, the experimentation period is July-September 2008, a period when additional experimental GPSRO data were available.

The OSE experiments that have run within this work package are as follows:

- *Control-full*: Full observing system (operational + additional GPSRO data);
- *Control-66*: 66% of GPSRO data coverage, rest of the observing system unchanged;
- *Control-33*: 33% of GPSRO data coverage, rest of the observing system unchanged;
- *Control-5*: 5% of GPSRO data coverage, rest of the observing system unchanged;
- *noAMDAR-66*: 66% of GPSRO data coverage, no AIRCRAFT data, rest of the observing system unchanged;
- *noAMDAR-33*: 33% of GPSRO data coverage, no AIRCRAFT data, rest of the observing system unchanged;
- *noRS-66*: 66% of GPSRO data coverage, no RADIOSONDE data, rest of the observing system unchanged;
- *noRSstrat-66*: 66% of GPSRO data coverage, no RADIOSONDE data above 50hPa, rest of the observing system unchanged;

- *noRS-33*: 33% of GPSRO data coverage, no RADIOSONDE data, rest of the observing system unchanged;
- *noRS-5*: 5% of GPSRO data coverage, no RADIOSONDE data, rest of the observing system unchanged;
- *noRSstrat-5*: 5% of GPSRO data coverage, no RADIOSONDE data above 50hPa, rest of the observing system unchanged;

Impact of aircraft denial

At first, the impact of assimilating aircraft data on the mean temperature analysis state is examined. Figure 1 shows the mean temperature analysis difference between aircraft denial and control experiments with 66% of the full GPSRO data coverage. On the left hand side, these differences are shown for 850 hPa (top) and 200 hPa (bottom) while the right-hand-side figure shows the zonal-vertical cross-section of the mean temperature analysis difference. Green-blue colours indicate areas where the aircraft data increase the mean temperature analysis. It can be seen that the warming effect of aircraft data is dominant, which suggests a warm bias of aircraft observations. The impact is largest in the 200-250 hPa layer and in the Northern hemisphere, where most of the flight level aircraft data are available for the assimilation.

At 200 hPa the mean temperature analysis difference over North-America, the North-Atlantic and Europe (i.e. in the vicinity of the most frequent flight routes) is betweeen 0.35-0.4 K, elsewhere it is smaller. At 850hPa, the biggest impact is over North-America, where the largest number of ascending and descending flights are found, and they result in a mean temperature analysis difference of 0.16 K.

A similar impact can be found when the analysis and first-guess fit to radiosonde temperatures is examined for the experiment with and without aircraft data. These are shown in Figure 2, where the red curves denote the *Control-66* experiment and the black ones the corresponding aircraft denial experiment. In terms of standard deviation of first-guess fit a very small improvement can be found around the main flight level, when aircraft data are present. However, the really large impact is seen in terms of biases, where the assimilation of aircraft data increases the analysis and first-guess bias with respect to radiosonde temperatures. This impact is the strongest at cca. 250 hPa over the Northern hemisphere and the North Pole area. The figure shows that in the 700-200 hPa layer the model analysis and first guess are warmer than radiosonde measurements (*observation-model* is negative) and this effect is amplified when aircraft data are present, in agreement with the warm aircraft biases suggested by Figure 1. Below 700 hPa, aircraft data still have a warming effect, but in this case the model has a cold bias against radiosondes and the warming effect of aircraft data improve the model fit to radiosondes.

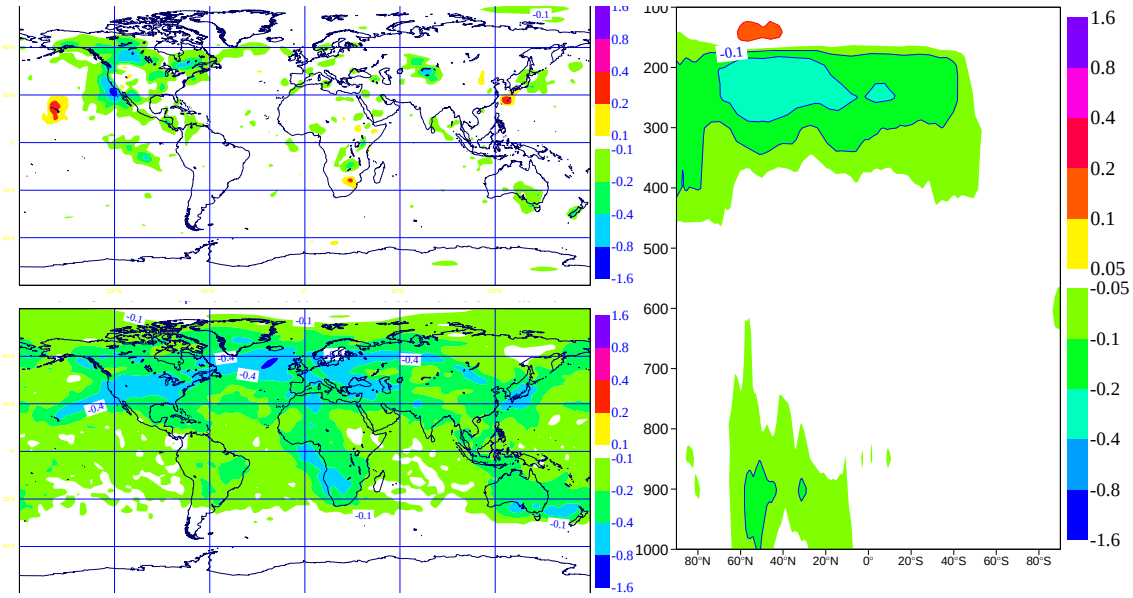


Figure 1: Mean temperature analysis difference (in K) between the aircraft denial and the control experiments for the 66% GPSRO data coverage case. On the left hand side difference fields for 850hPa (top) and 200hPa (bottom), on the right hand side zonal-vertical cross-section of the mean temperature analysis difference are displayed. Green-blue colours indicate areas where the aircraft data warm the analysis. Statistics computed from 00 and 12UTC analyses between 1 July - 30 September 2008.

The impact of aircraft data on wind analysis and short range forecast can be seen in Figure 3, which shows the analysis and first-guess fit to wind profiler data over North-America (left) and Europe (right). The two experiments are again compared to the *Control-66* experiment (red curves) and the corresponding aircraft denial (black curves). Near the tropopause the errors have a sharp maximum and, at the same time, aircraft data have the largest impact there because this is near the cruising level. The short range forecast (first-guess) fit in terms of standard deviation to both European and American profilers clearly improves when aircraft data are assimilated, even if the analysis fit is slightly deteriorated for American profilers.

Aircraft data also have an impact on the model's fit to satellite data. Figure 4 shows the GPSRO observation statistics for the Northern hemisphere derived from the same experiment-pair as shown in the previous figures. In the vicinity of the cruising level the aircraft data improve the fit to GPSRO observations both in terms of standard deviation and even more in terms of bias. The latter is somewhat surprising because this is the layer where the radiosonde fits indicated the largest bias increase. In the stratosphere, however aircraft data increase the bias with respect to GPSRO, as an integral effect of tropospheric biases. This can be explained by the shape of the GPSRO weighting function which peaks at the levels above and below the ray tangent point and which has a long flat tail below. As a consequence of this, when the GPSRO forward model is applied to a hypothetical temperature increment profile that is similar to the change of mean temperature state due to aircraft data assimilation the bending angle response profile is qualitatively similar to the difference between the red and black bias curves on Figure 4. It has to be noted that certain temperature perturbation profiles result in a close to zero response bending angle profile by the forward model, which means that temperature bias structures of such vertical shape remain unseen by GPSRO data. All this is well illustrated on Figure 5. The top panel shows a GPSRO weight function with the peak near the ray tangent point and with a long tail below. The bottom two panels show a temperature increment profile (left) that results in a quasi-zero bending angle response

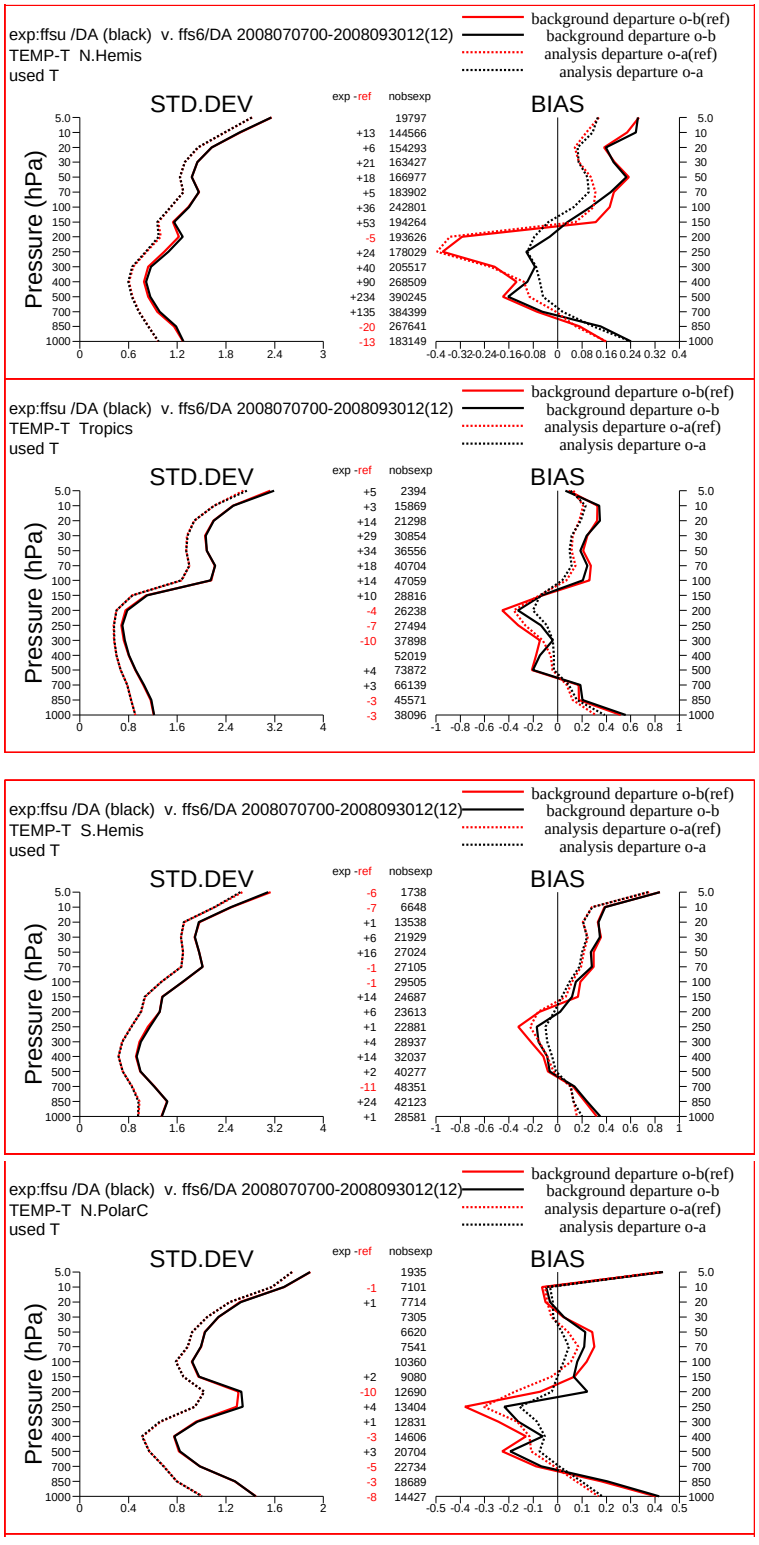


Figure 2: Fit of model first-guess (solid) and analysis (dashed) to radiosonde temperatures in the Northern hemisphere (top), Tropics (second from top), Southern hemisphere (third from top) and North Pole (bottom) from the Aircraft denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

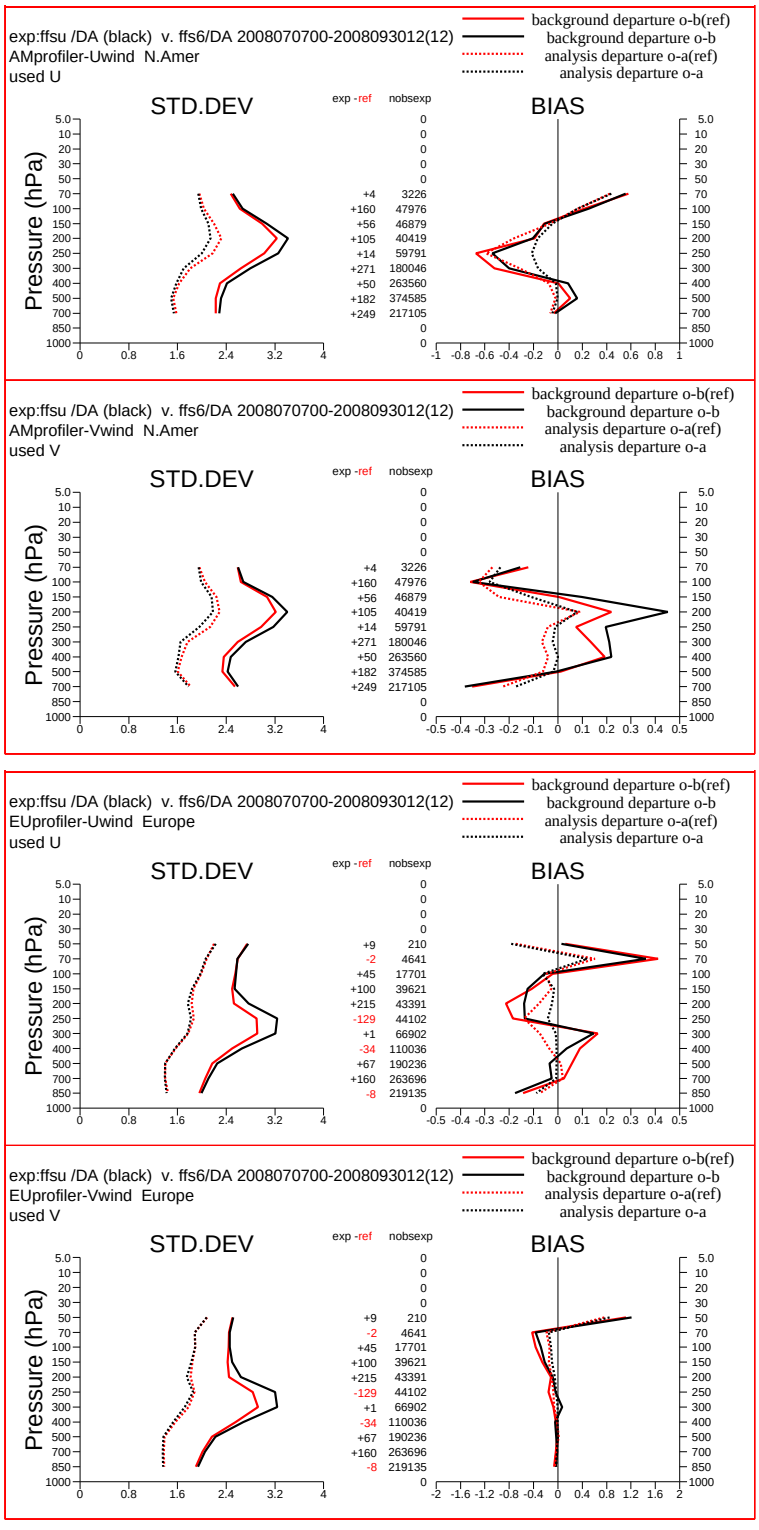


Figure 3: Fit of model first-guess (solid) and analysis (dashed) to American (top) and European (bottom) wind profiler measurements from the Aircraft denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

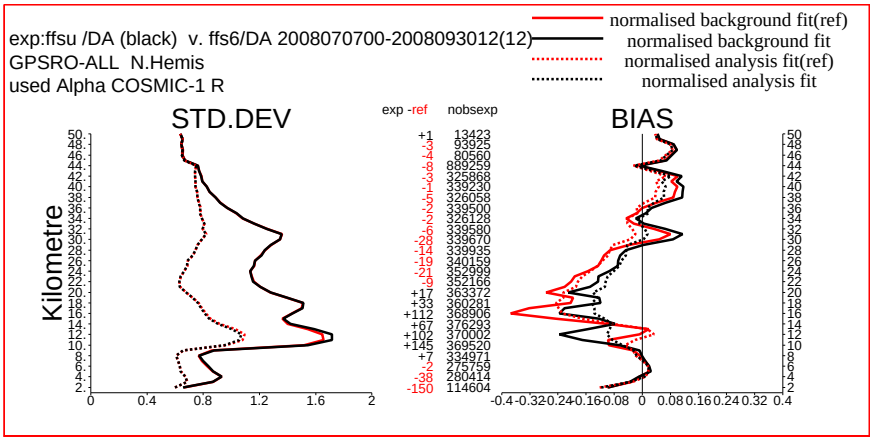


Figure 4: Fit of model first-guess (solid) and analysis (dashed) to GPSRO data over the Northern hemisphere from the Aircraft denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

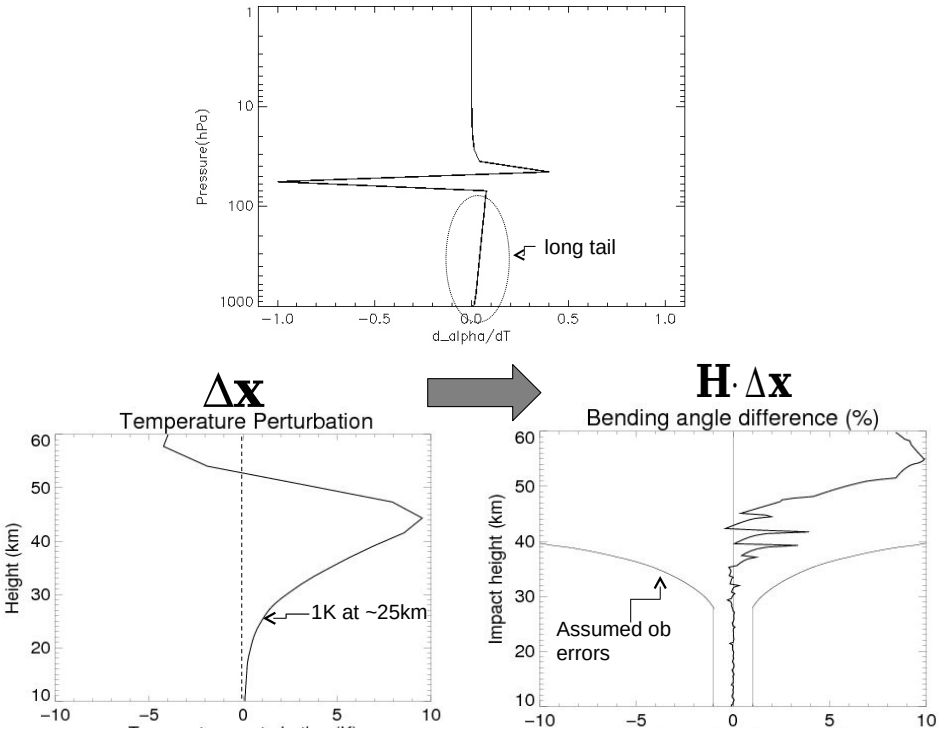


Figure 5: Sketch of a GPSRO weight function corresponding to a given ray tangent point (top) and illustration of the "null-space", i.e. a temperature increment profile (bottom left) that results in a quasi-zero bending angle response (bottom right). Figures taken from Sean Healey.

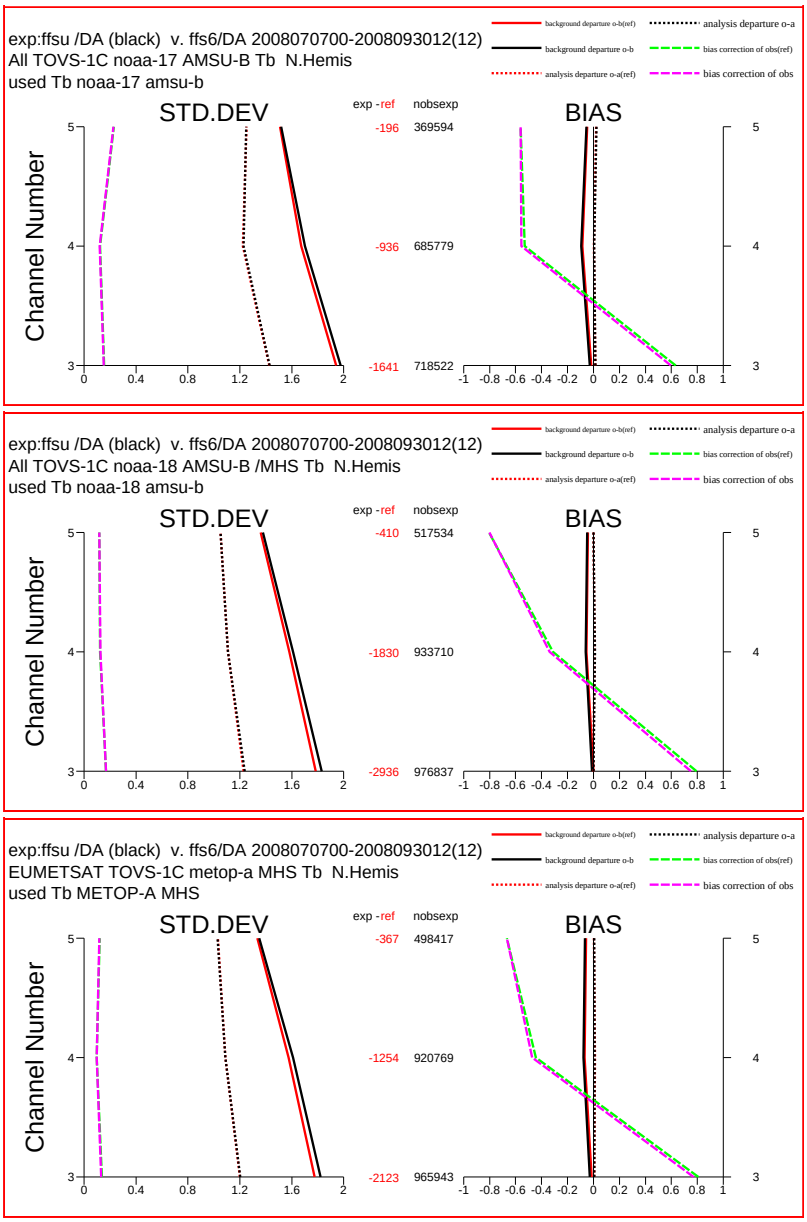


Figure 6: Fit of model first-guess (solid) and analysis (dashed) to NOAA-17 AMSU-B (top), NOAA-18 MHS (middle) and METOP MHS (bottom) radiances over the Northern hemisphere from the Aircraft denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

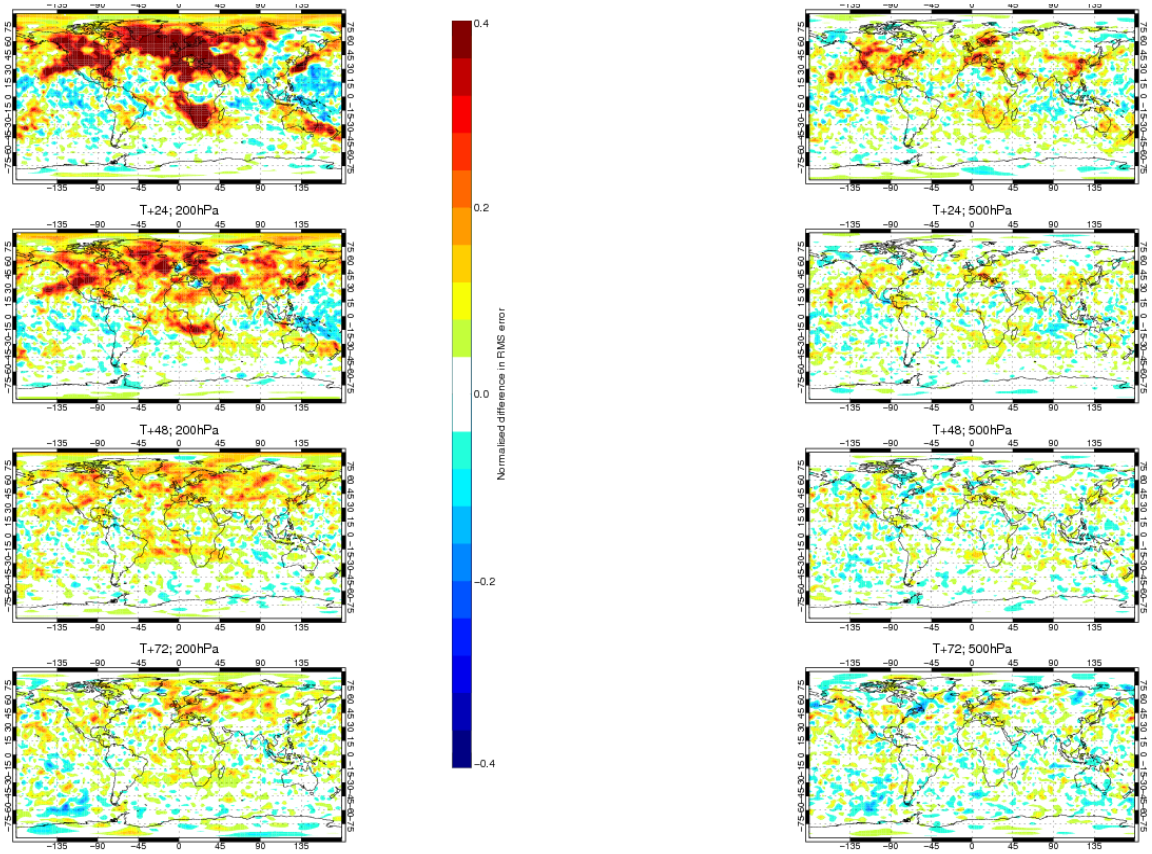


Figure 7: Normalized RMS forecast error difference between Aircraft denial and Reference experiments with 66% GPSRO data coverage. Maps on the left show 200 hPa temperature, on the right 500 hPa RMS error differences. Positive values (red colours) indicate positive impact of the assimilated aircraft data. Panels show forecast ranges of 12, 24, 48 and 72 hours (from top to bottom). Forecast verification is against operational analyses; verification period is 7 July - 29 September 2008.

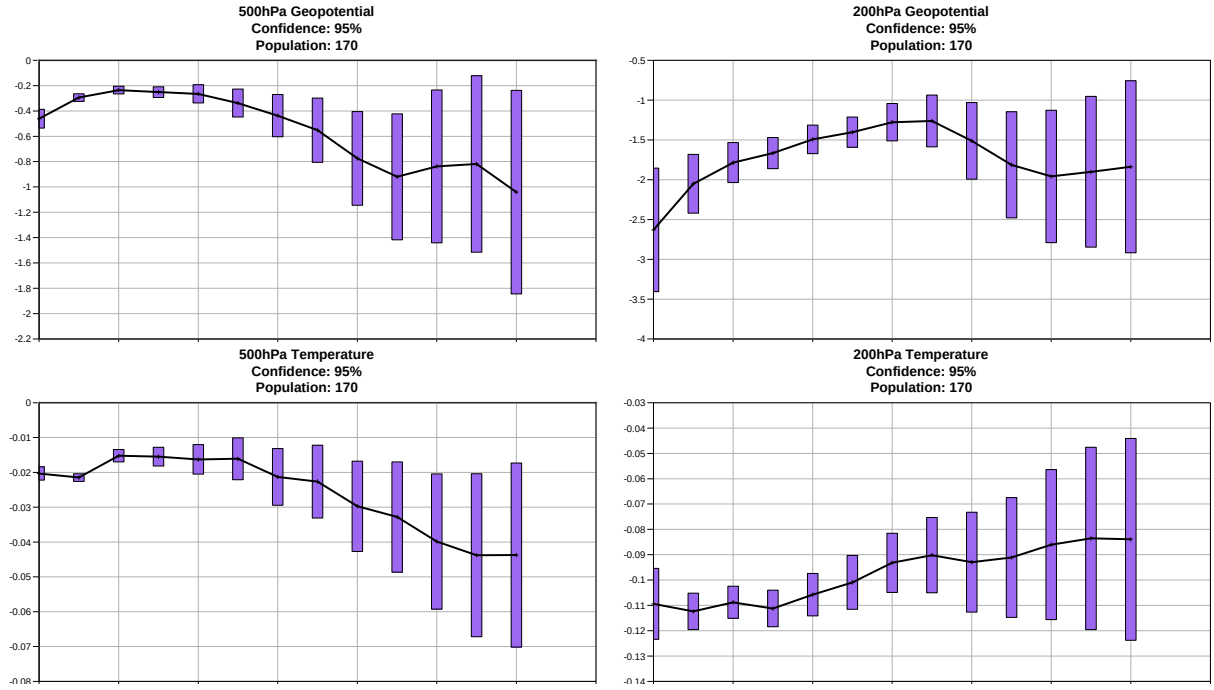


Figure 8: Time evolution of RMS forecast error difference between Reference and aircraft denial experiments with 66% GPSRO data coverage over the Northern hemisphere. Variables and levels are as indicated on the figure. Forecast verification is against operational analyses; verification period is 7 July - 29 September 2008.

(right) in a fairly deep layer of the atmosphere (although there is a non-zero response very high in the atmosphere, this response can not constrain the analysis due to the large assumed observation errors as also seen on the right hand panel). This also means that using GPSRO data as the only temperature analysis anchors is unsafe, even if the denial experiments show a good anchoring effect for a given type of model biases.

As far as the impact of aircraft data assimilation on satellite radiance fits is concerned the largest (positive) impact can be identified, again somewhat surprisingly, on the humidity-sensitive AMSU-B and MHS radiance data fits, especially for channel-3 that peaks in the upper troposphere. This can be seen in Figure 6 together with a moderate sensitivity of the mean value of the variational bias correction to the use of aircraft data. Even if aircraft observations do not contain direct humidity information, the background error model generates humidity increments due to temeprature observations, that can be rather large, especially near saturation. Indeed, the mean humidity analysis state shows differences due to aircraft data assimilation, especially in the upper troposphere and the root-mean-square difference of analyses with and without aircraft data also reachs out as much as 30% of the humidity value itself at certain regions of the atmosphere (not shown here).

The impact of aircraft data on the model forecast performance is shown in Figure 7, where the normalized root-mean square (RMS) temperature forecast error reduction maps are showm for 200 hPa (left) and 500 hPa (right). The panels show forecast ranges of 12, 24, 48 and 72 hours from top to bottom. Red colours indicate error reduction. The RMS error reduction is clearly stronger at 200 hPa and over the Northern hemisphere, where the highest aircraft data density is available. The time evolution of temperature and geopotential height RMS error difference for 200 and 500hPa over the whole Northern hemisphere can be seen on Figure 8 and it confirms the strong impact of aircraft data for the whole 6 day forecast time

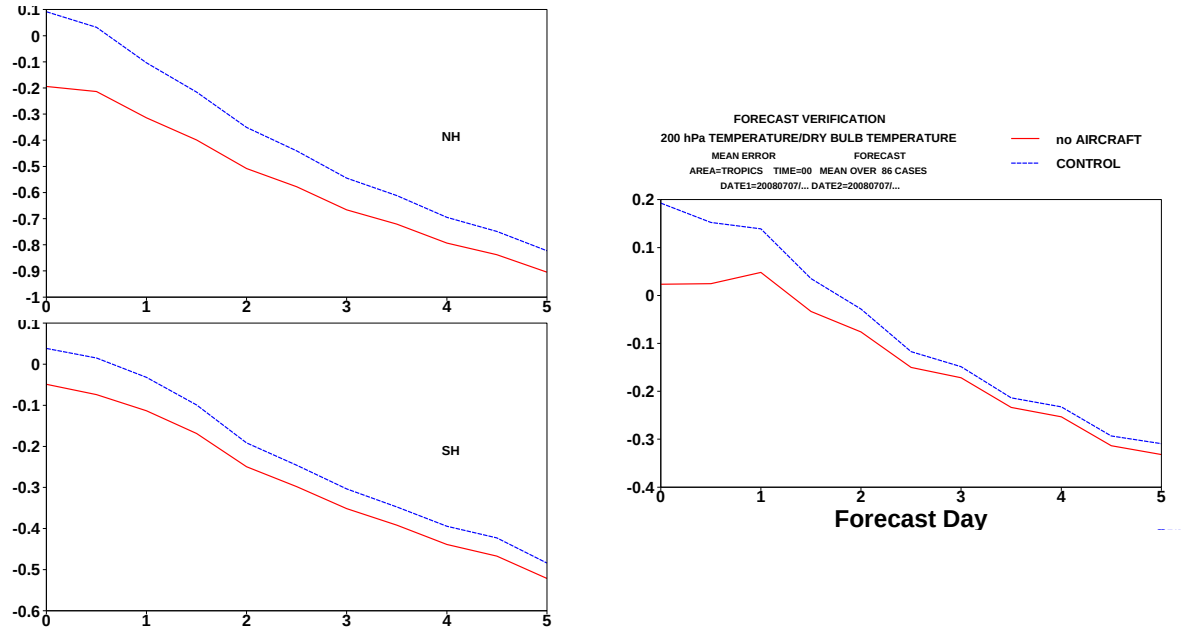


Figure 9: 200 hPa temperature forecast bias for the Northern hemisphere (top left), Southern hemisphere (bottom left) and Tropics (right) for the Reference (blue curves) and Aircraft denial (red curves) experiments with 66% GPSRO data coverage. Verification is against operational analysis, verification period is 7 July - 29 Septembe 2008.

range.

Wind forecast impacts are very similar (not shown here). The forecast bias impact is in agreement with what has been found in the comparison of analysis and first-guess fits to other observations. This is shown in Figure 9 for 200 hPa temperature in the Northern hemisphere (top left), Southern hemisphere (bottom left) and Tropics (right) which were verified with the operational analysis. Clearly the assimilation of aircraft data warms the analysis and this impact is persistent along the forecast until day-5. The impact is the largest over the Northern hemisphere where it is 0.3 K at initial time and then gradually reduces to 0.1 K at day-5.

Impact of radiosonde data denial

The mean temperature analysis impact of radiosonde observations is summarized in Figure 10. This result was derived again from the experiments with 66% GPSRO data coverage. The left panel shows the analysis and first-guess bias against radiosonde temperature observations for Reference over the Northern hemisphere. The model is systematically warmer than radiosonde observations between 700 hPa and 150 hPa and colder in the boundary layer and in the stratosphere. The basic structure is similar for the Southern hemisphere and for the Tropics with the difference that the lower tropospheric layer of cold model bias is deeper there (not shown).

The middle panel shows the mean temperature analysis difference between the Radiosonde denial and the Reference experiments at 500 hPa (top) and 850 hPa (bottom). These are in agreement with radiosonde analysis departure statistics, i.e. assimilating radiosonde data makes the analysis warmer at 850 hPa, especially over land. On the other hand, the opposite effect can be observed over the Northern hemisphere 500 hPa, while over the Southern hemisphere the mean temperature impact is rather neutral. The right panel shows the same results as a zonal-vertical cross section. It has to be mentioned that in the 200-300 hPa layer of the cross section, where the left figure showed the biggest anomaly, no systematic impact of radiosonde denial can be observed. This suggests that the temperature model bias with respect to radiosondes at this layer is fully dominated by the large number of aircraft flight level data so that adding or removing radiosonde data hardly affects the aircraft data impact.

The assimilation of radiosonde data has an impact on the mean humidity analysis state as well. Figure 11 shows this impact as the mean total column water vapour (TCWV) difference field (left) and the mean zonal-vertical specific humidity difference cross-section (right) between the Radiosonde denial and Reference experiments in the case of 66% GPSRO data coverage. The TCWV-difference map shows that radiosondes moisten the analysis over most of the continental areas with some exceptions, e.g. over China, the North of the Black-Sea or the Amazon Valley, where only few sondes are available. The vertical cross-section of specific humidity shows that this overall moistening impact is originating from the lower troposphere, while between 200 and 50 0hPa, especially over the Northern hemisphere, radiosondes rather dry the mean analysis state.

The largest impact of radiosonde data assimilation on the mean analysis and first-guess fit to satellite data can be seen on HIRS channel-11 and on the ten assimilated IASI humidity channels, as shown in Figure 12. This impact is stronger for first guess biases and these seem to be related to the upper tropospheric drying as a consequence of radiosonde observation assimilation. The impact of radiosonde assimilation on the mean bias correction of satellite radiances is surprisingly small, and for most satellites/channels completely undetectable.

As opposed to the aircraft data impact, the assimilation of radiosonde observations does not change the analysis and first guess bias characteristics with respect to GPSRO data (not shown here). In terms of

standard deviation, however, there is a small, but clear improvement of the first-guess fit to GPSRO data when radiosondes are assimilated and this improvement extends over the entire stratosphere as seen on the left panel of Figure 13. The right panel shows the same GPSRO statistics, but for the experiment where only stratospheric radiosonde data have been rejected (data rejection above 50hPa) and here no improvement can be detected. This clearly proves that although the impact in terms of GPSRO bending angles is the strongest in the higher stratosphere, this impact is coming from radiosonde observations below 50hPa. The explanation is the same as described above for aircraft denial impacts: the weighting function of GPSRO bending angles spreads well below the ray tangent point, thus stratospheric GPSRO bending angles show a strong sensitivity to tropospheric temperature profiles (see again Figure 5 and the corresponding explanation).

The assimilation of radiosondes has, as expected, also a large impact on the forecast scores. This is illustrated in Figure 14, Figure 15 and Figure 16 which show the zonal-vertical cross-sections of normalized RMS forecast error reduction (positive values mean error reduction) due to the assimilation of radiosonde data for temperature , wind and relative humidity, respectively. These results have been derived from the experiments with 66% GPSRO data usage and the verification was performed against the operational analyses. The cross-hatched areas indicate areas where the RMSE differences are significant at the 95% confidence level. It can be seen that the error reduction is dominant, especially over the Northern hemisphere, and the significant positive impact lasts up to 2-3 days for humidity and up to 4-5 days for temperature and wind.

The above experiments thus produced an overall positive forecast impact both for radiosonde and aircraft observation assimilation. Figure 17 compares their respective impact by calculating the normalized forecast RMS error difference between radiosonde denial and aircraft denial experiments, both with 66% GPSRO data usage. The figure shows zonal-vertical cross-sections of geopotential height RMSE difference (left) and RMSE difference maps of 500 hPa geopotential height forecasts (right), again verified against operational analyses.

Blue areas indicate regions where aicraft data are more beneficial in terms of RMSE scores. From the zonal-vertical cross sections it is clearly seen that near the cruising altitude aircraft data contributes much more to forecast quality than radiosondes, especially over the Northern hemisphere. This dominance extends down to 400-500 hPa. On the other hand, in the Northern hemispheric extra-tropical lower troposphere the opposite can be observed and radiosondes are more beneficial. At 500 hPa geopotential height, the scores over North-America, Europe and much of the Northern Atlantic show a dominant positive impact from aircraft observations, while the continental areas of Asia, Africa, South-America indicate a radiosonde data dominance, at least in the early forecast hours.

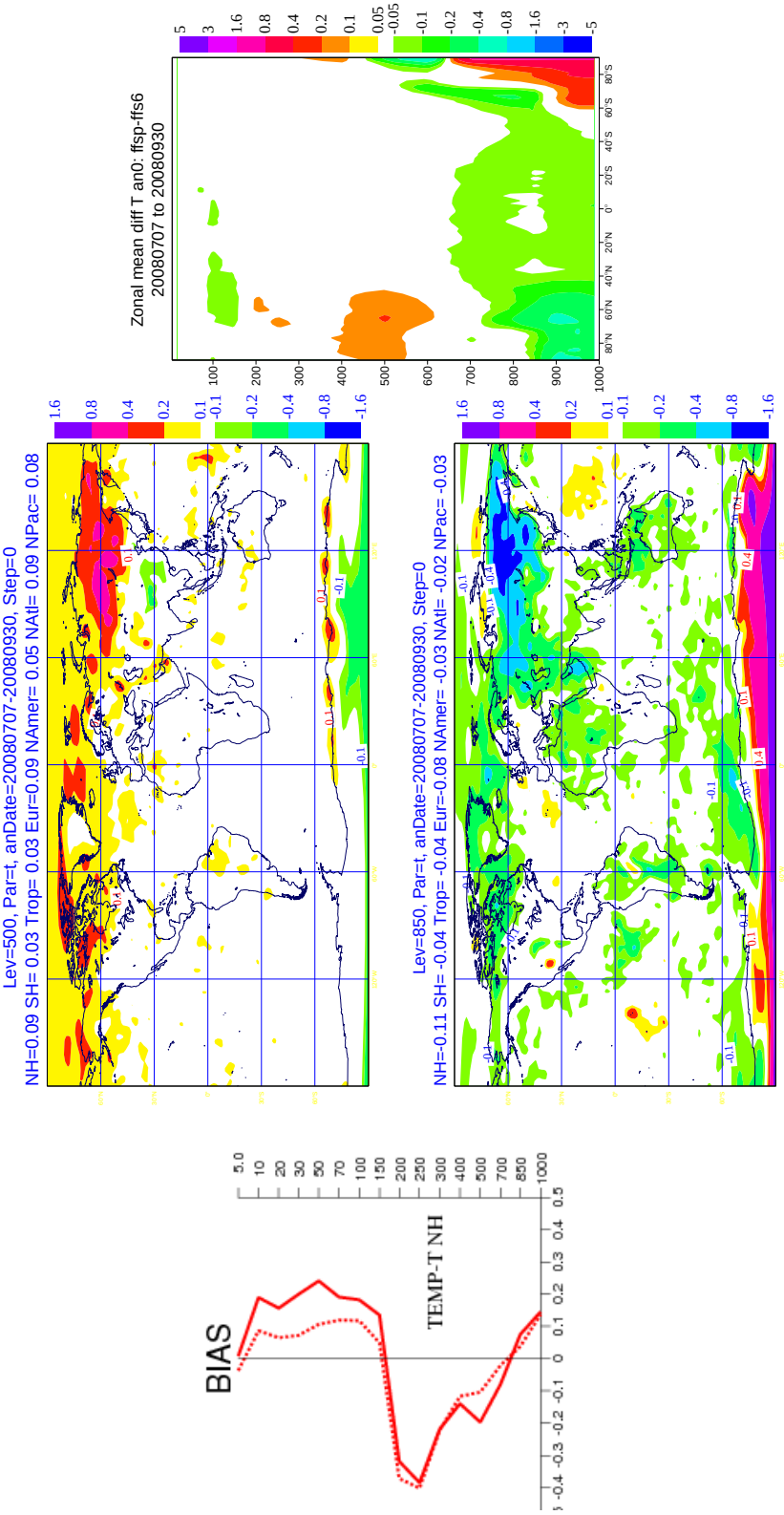


Figure 10: Mean temperature analysis impact derived from the analyses between 07/07-29/09/2008. Left panel: Bias of model first-guess (solid) and analysis (dashed) against radiosonde temperatures over the Northern hemisphere for the Reference experiment with 66% GPSRO data. Middle panel: Mean temperature analysis difference between the Radiosonde denial and the Reference experiment with 66% GPSRO data on 500hPa (top) and 850hPa (bottom). Right panel: Zonal-vertical cross-section of Mean temperature analysis difference between the Radiosonde denial and the Reference experiment with 66% GPSRO data.

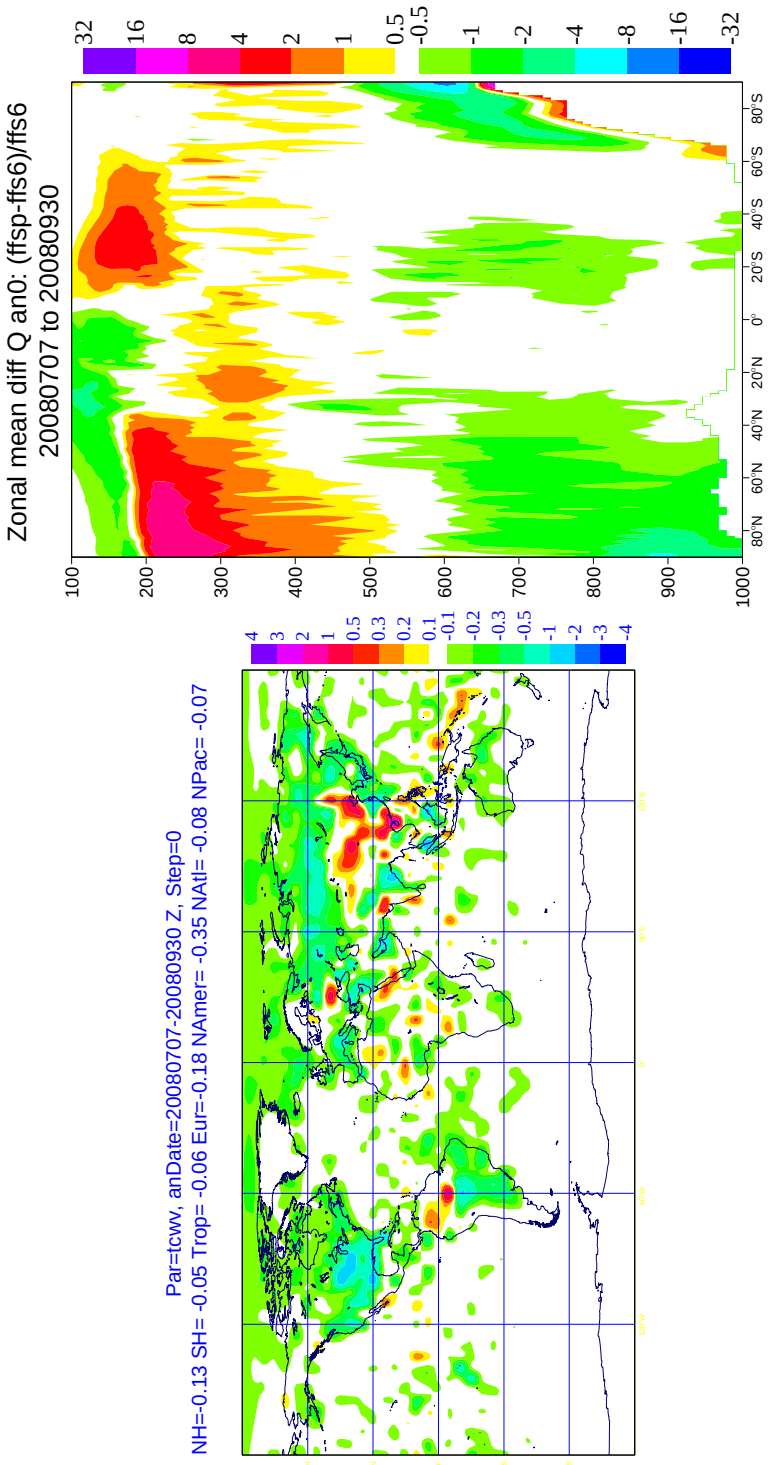


Figure 11: Mean total column water vapour (TCWV) difference field (left) in kgm^{-2} and the mean zonal-vertical specific humidity normalized difference cross-section (right) in % between the Radiosonde denial and Reference experiments with 66% GPSRO data coverage. Statistics derived from the analyses between 7 July and 29 September 2008.

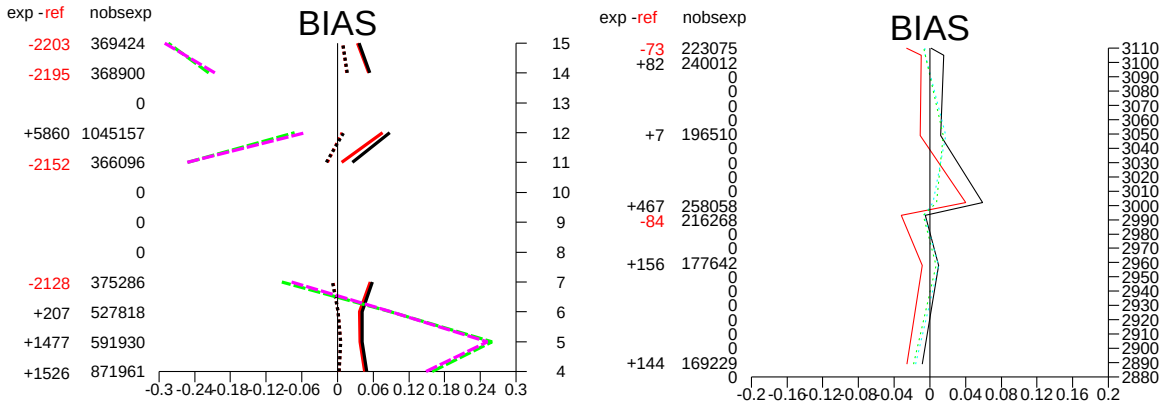


Figure 12: Bias of model first-guess (solid) and analysis (dashed) against METOP HIRS (left) and METOP IASI radiances over the Northern hemisphere from the Radiosonde denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

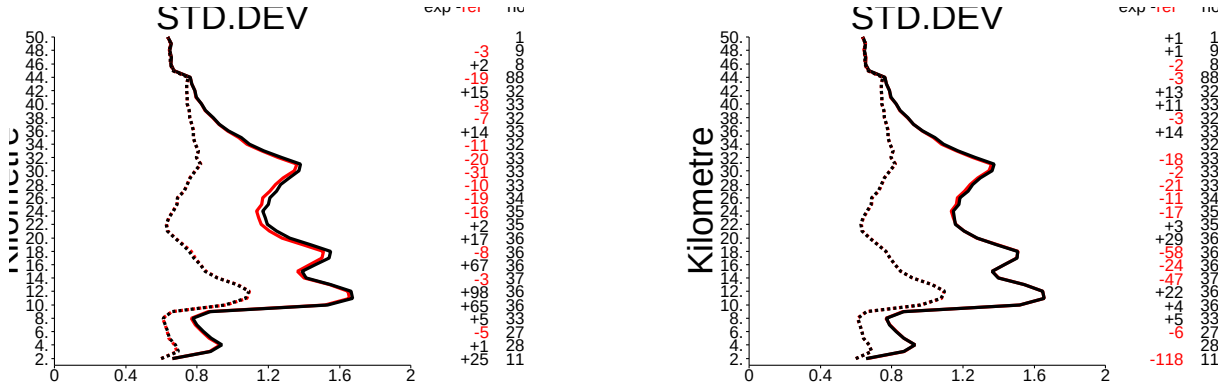


Figure 13: Standard deviation of model first-guess (solid) and analysis (dashed) fit to GPSRO bending angles over the Northern hemisphere from the total Radiosonde denial (black) and Reference (red) experiments with 66% GPSRO data coverage. Left panel shows results with total radiosonde denial and right panel shows results with stratospheric radiosonde denial, i.e. where all radiosonde data above 50hPa have been rejected. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

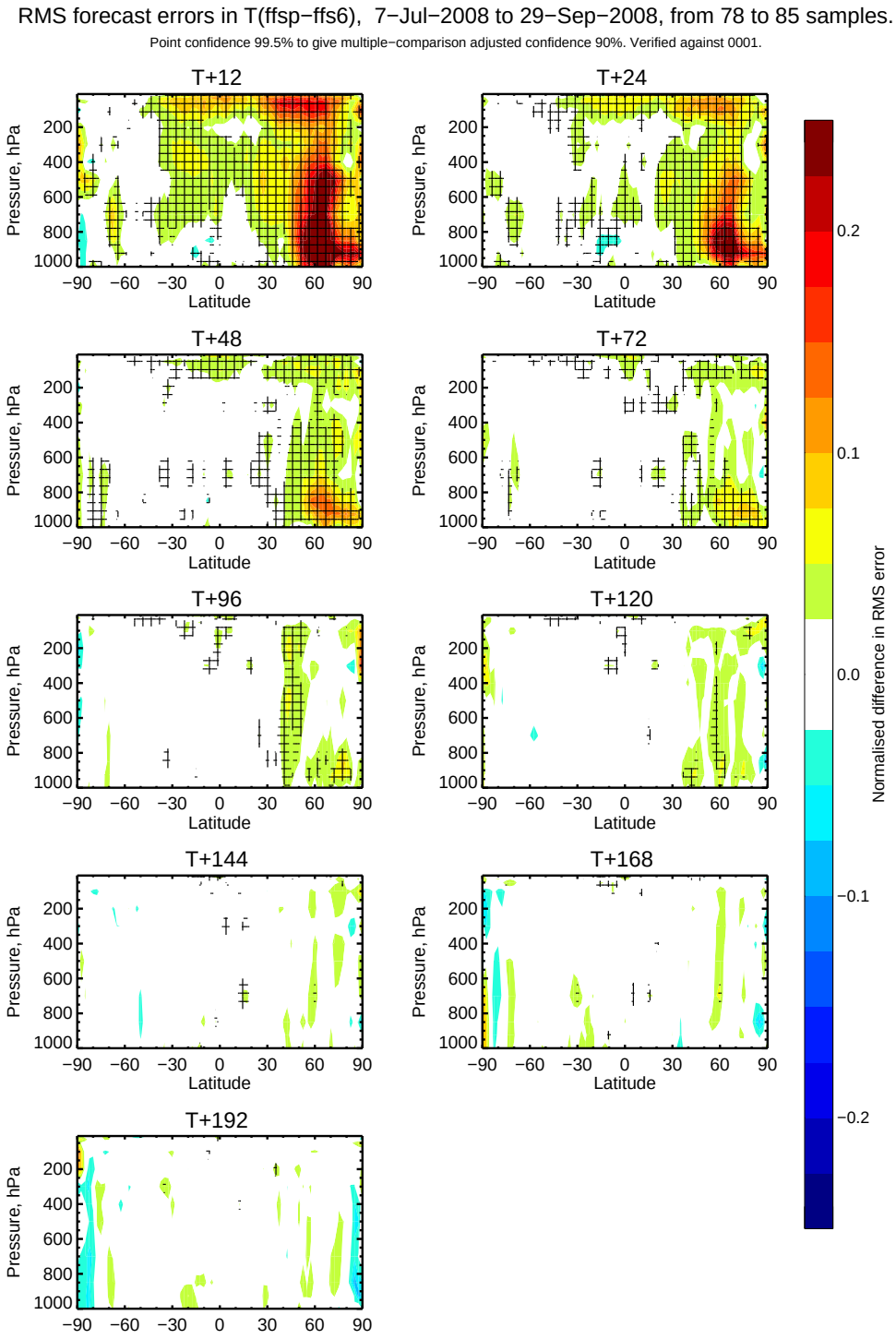


Figure 14: Zonal-vertical cross sections of normalized RMS forecast error reduction (positive values mean error reduction) due to assimilation of radiosonde data for temperature. Results were derived from the experiments with 66% data GPSRO data coverage and verified against operational analyses. The crosses indicate areas where the RMSE differences are significant at 95% confidence level. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

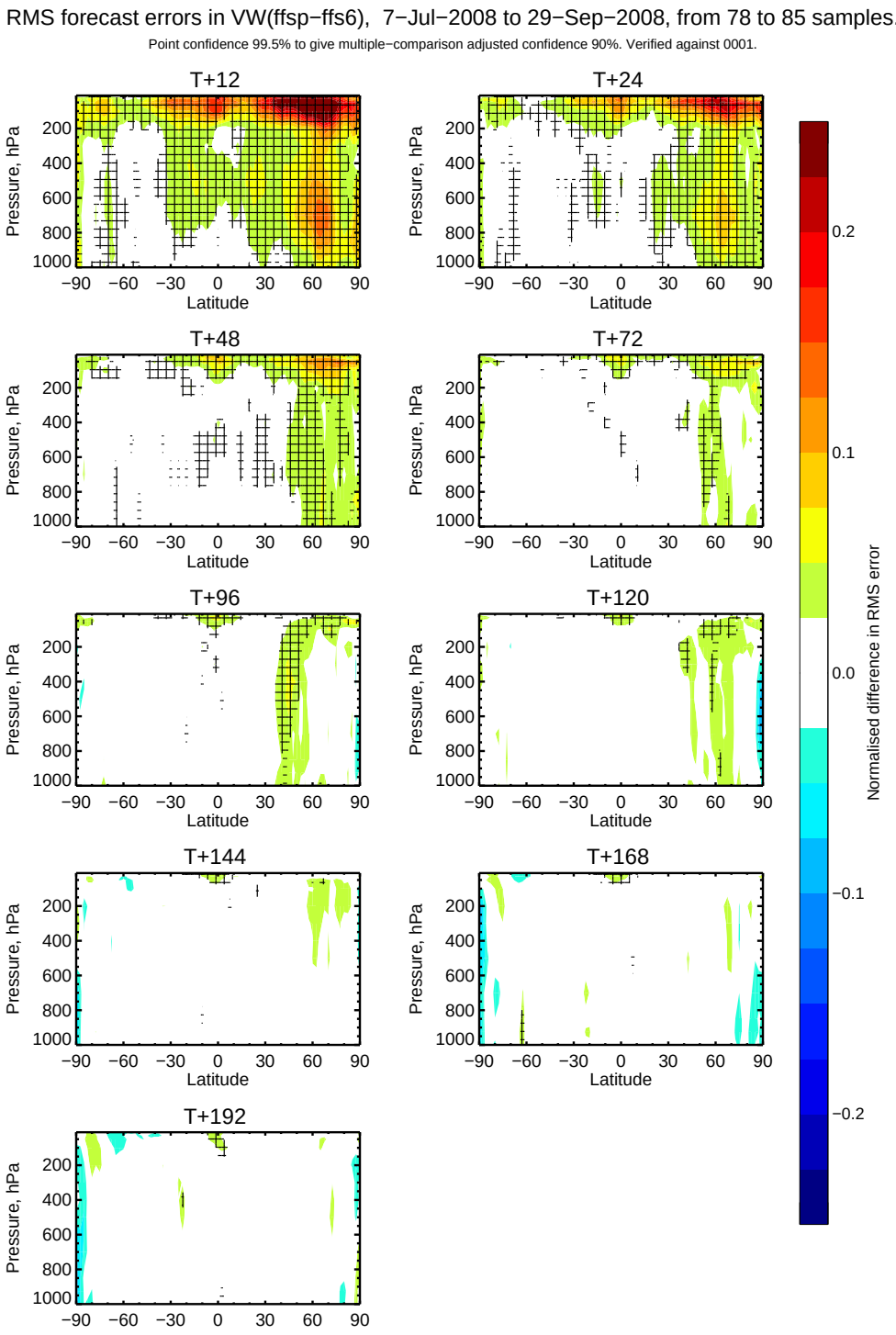


Figure 15: Same as Figure 14 but for vector wind.

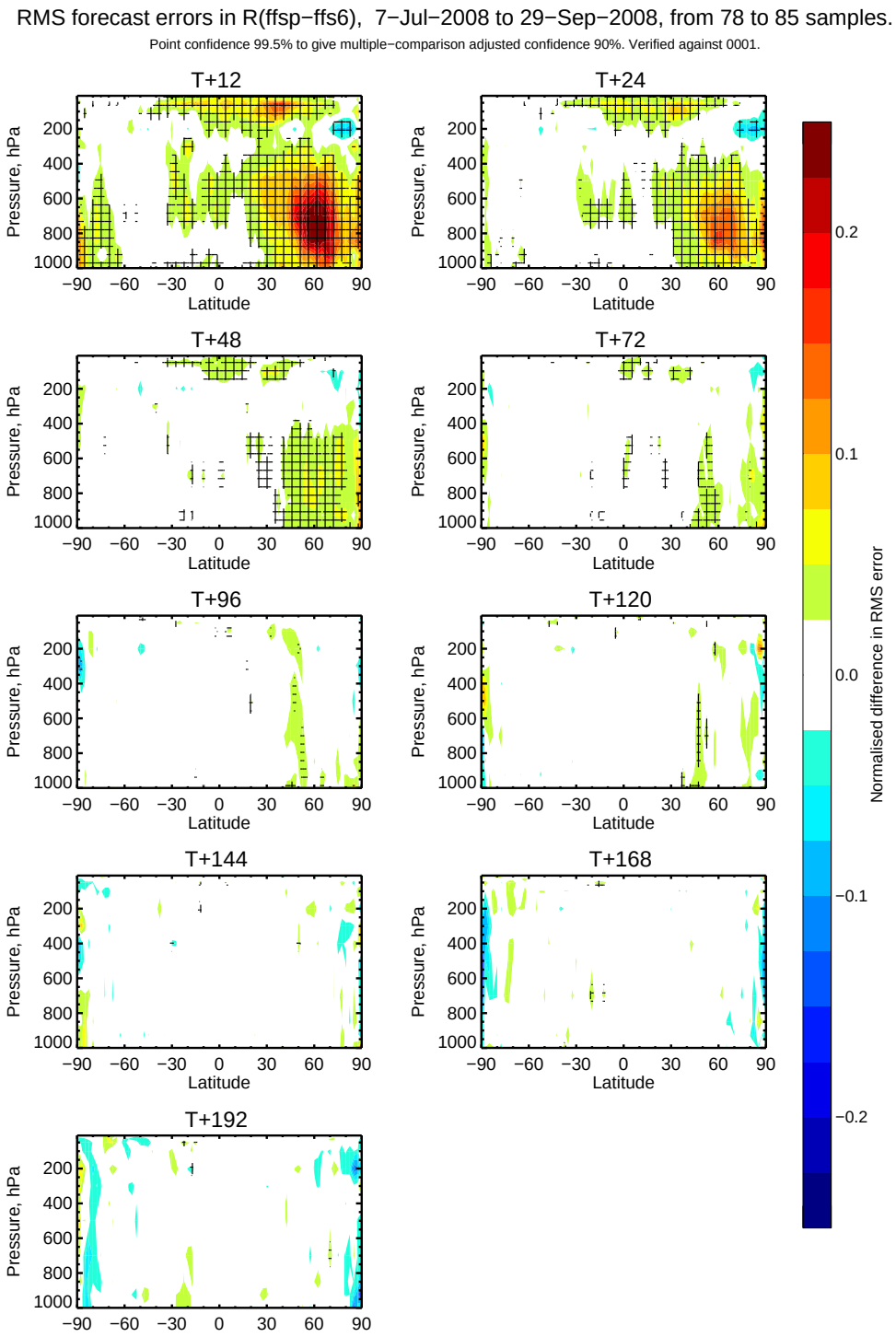


Figure 16: Same as Figure 14 but for relative humidity.

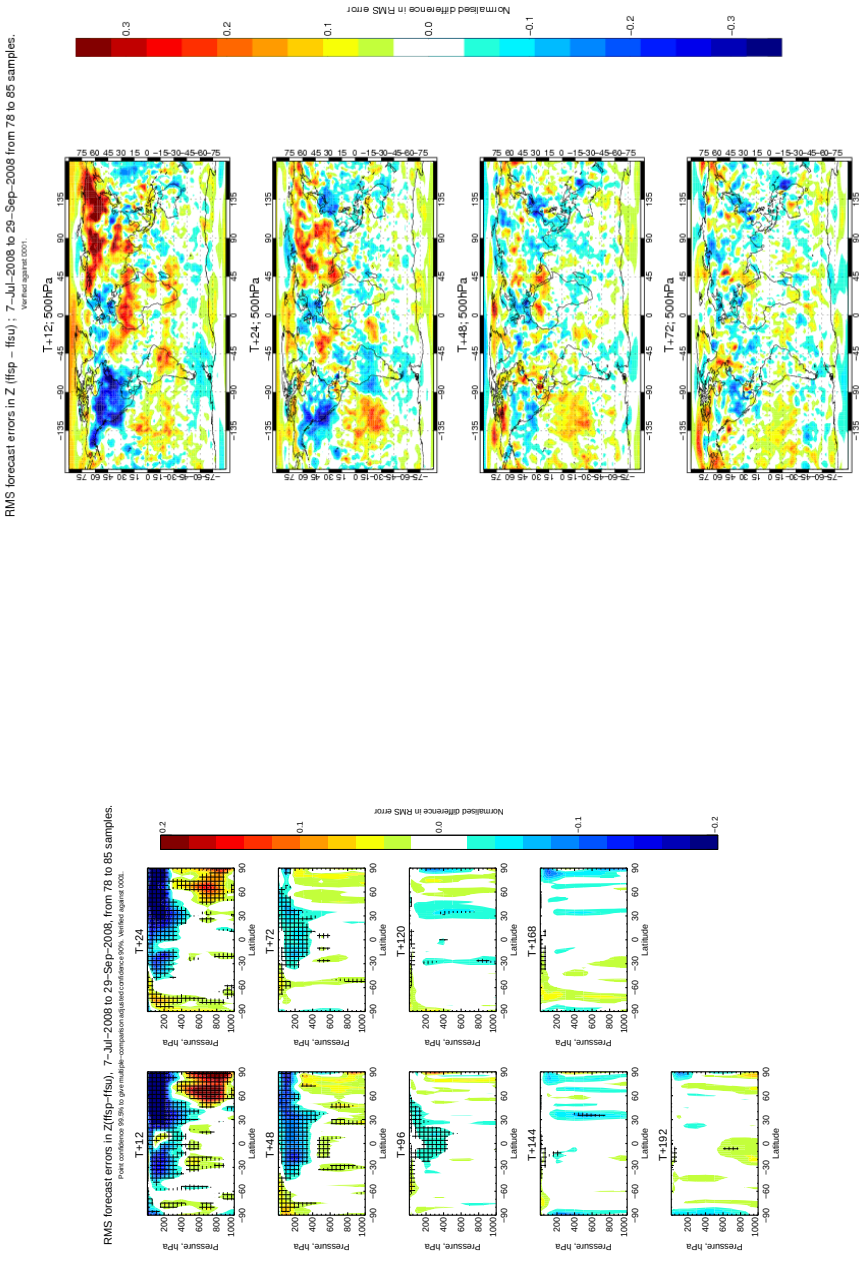


Figure 17: Normalized RMS forecast error difference between the Radiosonde denial and Aircraft denial experiments for zonal-vertical cross-section of geopotential height RMSE difference (left) and RMSE difference map of 500 hPa geopotential (right). Blue areas indicate areas, where aicraft data are more beneficial in terms of RMSE scores. Results were derived from the experiments with 66% GPSRO data coverage and verified against operational analyses. The crosses indicate areas where the RMSE differences are significant on 95% confidence level. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

Impact of GPSRO data thinning

So far, several aspects of the aircraft and radiosonde data impact on analyses and forecasts have been examined assuming 66% of the totally available GPSRO data coverage. As mentioned earlier, the 3-month period used for the experiments was chosen to coincide with the maximum of the historically available GPSRO constellation in summer 2008. Today's constellation is degraded by about 1/3 so that 66% of the summer 2008 data numbers are usable today. Earlier studies (e.g. Radnoti et al. 2010) have discussed how GPSRO data constrain temperature profiles and surface pressure in the analysis and showed that this data type is complementary to conventional observations.

Only selected examples of the GPSRO data impact on satellite radiance bias corrections are shown to illustrate the anchoring mechanism. Figure 18 shows the mean temperature analysis difference at 4 different vertical levels between two reference experiments, one with 33% and the other with 66% GPSRO data usage. Green and blue areas denote where the additional GPSRO data increase mean temperatures in the analysis, while yellow and red areas are where the data reduce temperatures. It is seen that in the stratosphere GPSRO data have a warming effect while in the troposphere they cool the analysis, which is in agreement with known model biases that are in contradiction with the data. The resulting analysis is also closer to radiosonde observations everywhere except for the lower boundary layer, where the model is generally colder than radiosondes (Figure 10) and GPSRO data reduce temperatures further. GPSRO data have very little sensitivity at these levels and the cooling originates from propagating the upper level increments to the surface through the operator.

At tropospheric levels the GPSRO data impact is much smaller over continents than over oceans. This is especially true for Europe and North America and one obvious reason is that the analysis is more constrained by conventional observations and therefore the GPSRO density has a less important role. It is worth to take a deeper look at this issue, for example at 200 hPa, where the dominant impact of aircraft data at cruising level was seen. Figure 19 displays the same behaviour as Figure 18, but only at 200 hPa with both radiosonde and aircraft data assimilated (left), without radiosonde data (top right) and without aircraft data (bottom right).

If aircraft data are missing the overall contrast is smaller, i.e. the aircraft impact is the primary source of the systematic analysis difference that GPSRO data assimilation is compensating for. Secondly, if radiosonde data are missing the land-sea contrast disappears except over the US. This means that in the absence of radiosondes GPSRO data seem to assume the same role in terms of correcting systematic errors over land and over sea, except for the US, where aircraft data are dominating the analysis. The contrast due to GPSRO density suddenly drops at 200hPa at the Southern-Hemisphere mid-latitudes. This is true in every combination of radiosonde and aircraft data usage. To verify this Figure 20 shows the zonal-vertical cross-section of mean temperature analysis difference between the 33% and 66% GPSRO usage scenarios in case of radiosonde denial (left) and aircraft denial (right). The opposite impact in the troposphere and stratosphere is clearly visible and in the Southern mid-latitudes indeed there is a relatively deep neutral layer below the Tropopause. It is also interesting to see the dipole structure in the mean temperature impact in the lower stratosphere over the Tropics.

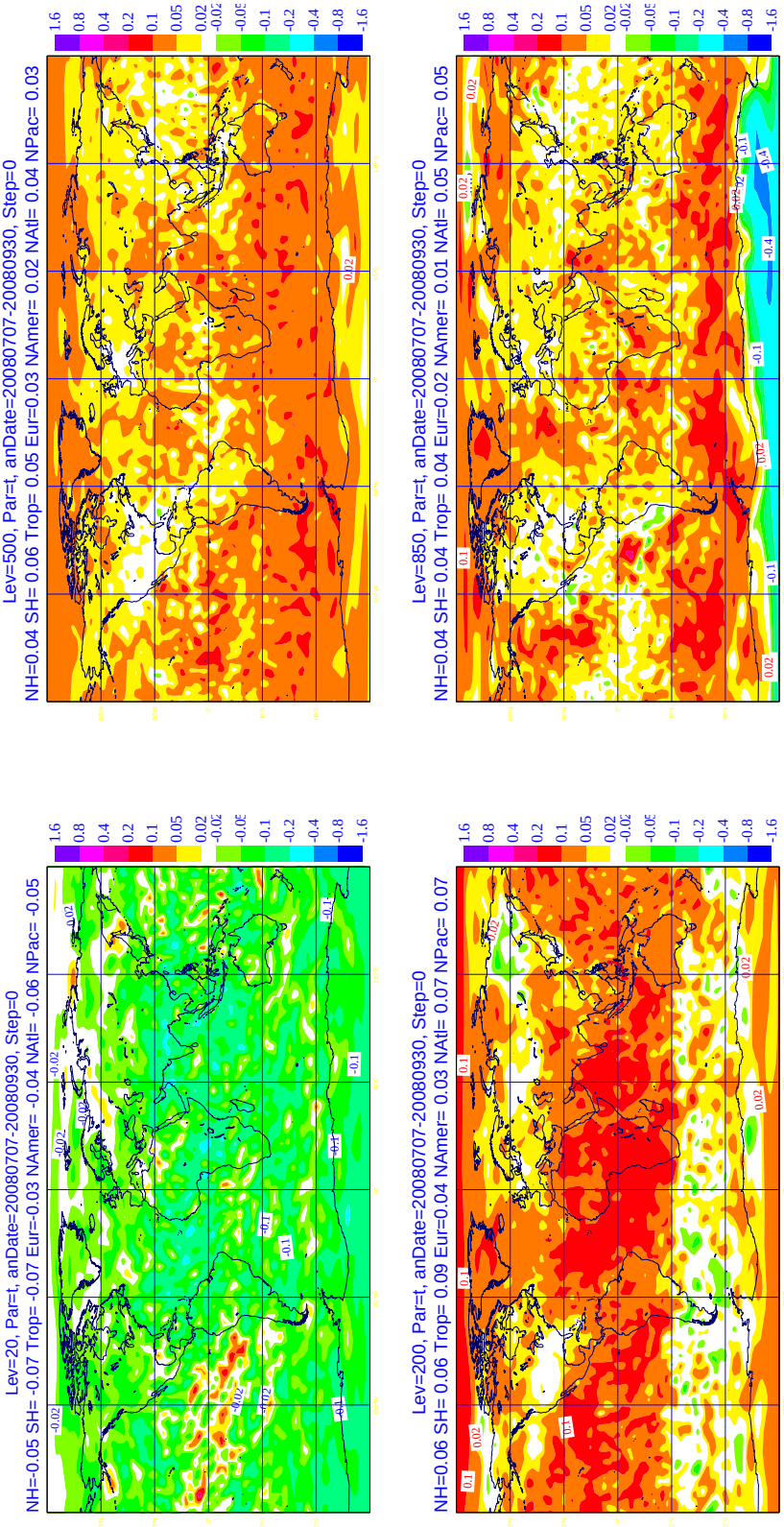


Figure 18: Mean temperature analysis difference at different vertical levels (top left: 20 hPa, bottom left: 200 hPa, top right: 500 hPa, bottom right: 850 hPa) between two experiments with 33% and 66% GPSRO data usage. Green and blue areas denote where the additional GPSRO data increase temperature in the analysis, yellow and red areas denote where they reduce temperatures. Statistics were computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

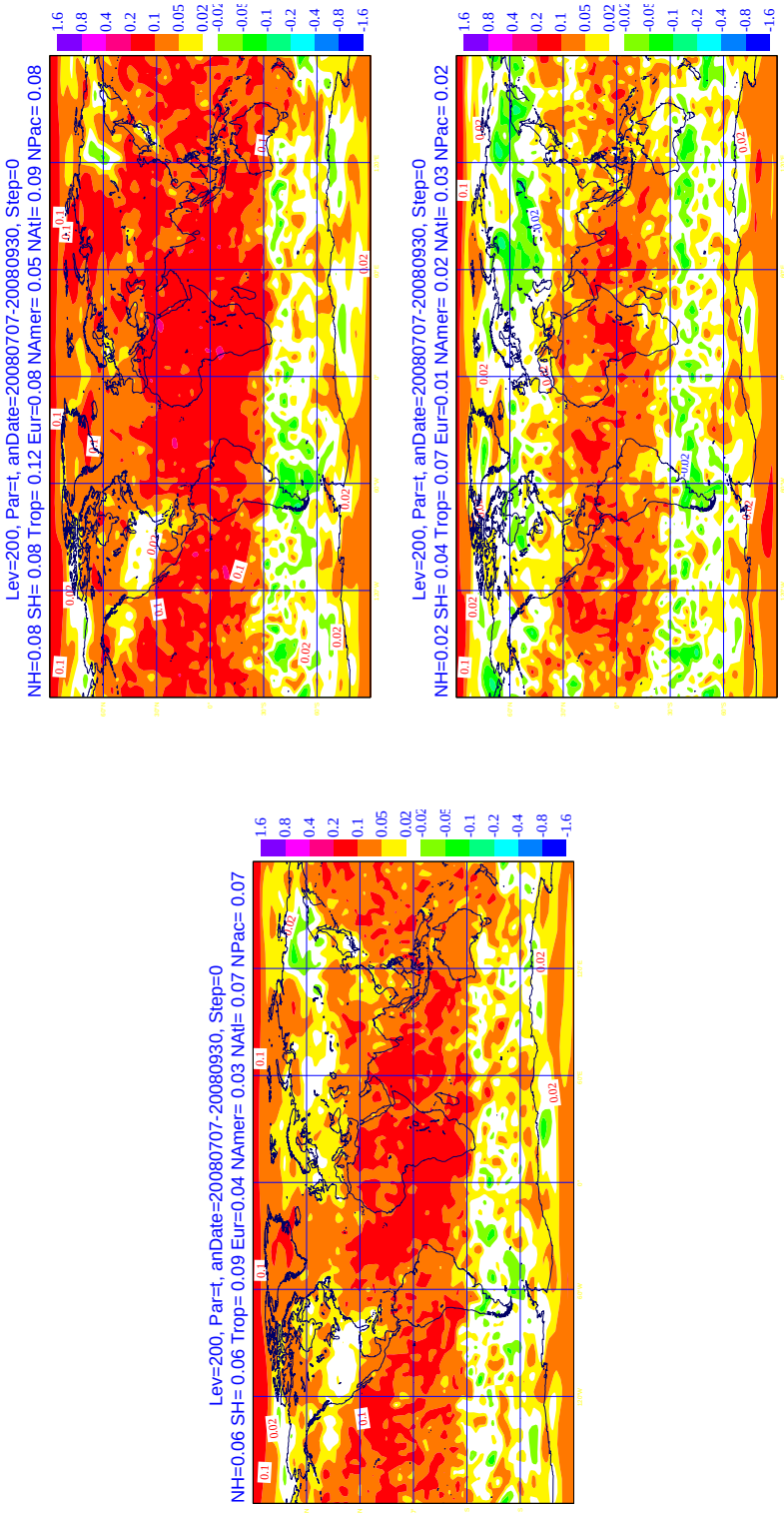


Figure 19: Mean temperature analysis difference at 200hPa (left: both radiosonde and aircraft data assimilated, top right: radiosondes not assimilated, bottom right: aircraft data not assimilated) between two experiments with 33% and 66% GPSRO data usage. Green and blue areas denote where the additional GPSRO data increase temperature in the analysis, yellow and red areas denote where they reduce temperatures. Statistics were computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

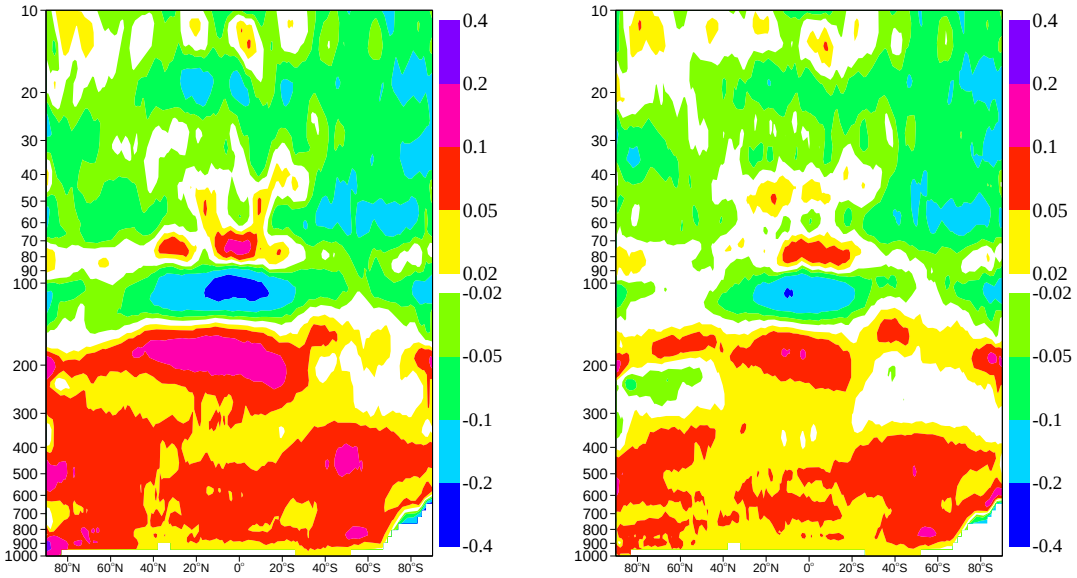


Figure 20: Zonal-vertical cross section of mean temperature analysis difference (left: radiosonde data not assimilated, right: aircraft data not assimilated) between two experiments with 33% and 66% GPSRO data usage. Green and blue areas denote where the additional GPSRO data increase temperature in the analysis, yellow and red areas denote where they reduce temperatures. Statistics were computed from 00UTC analyses between 7 July and 30 September 2008.

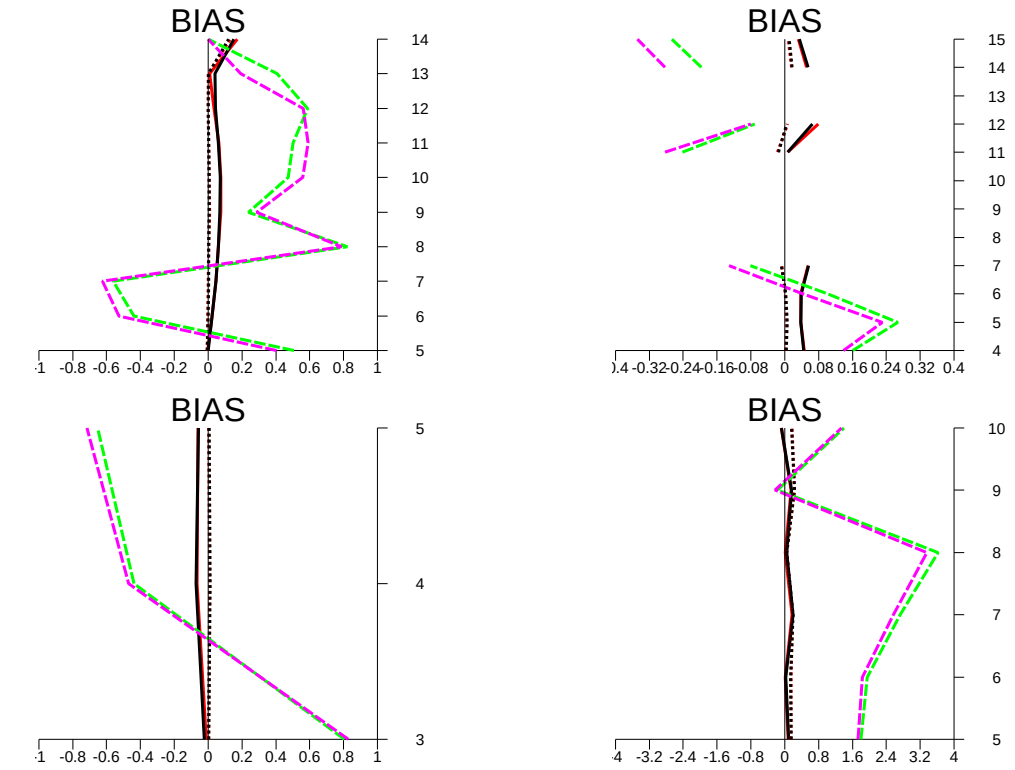
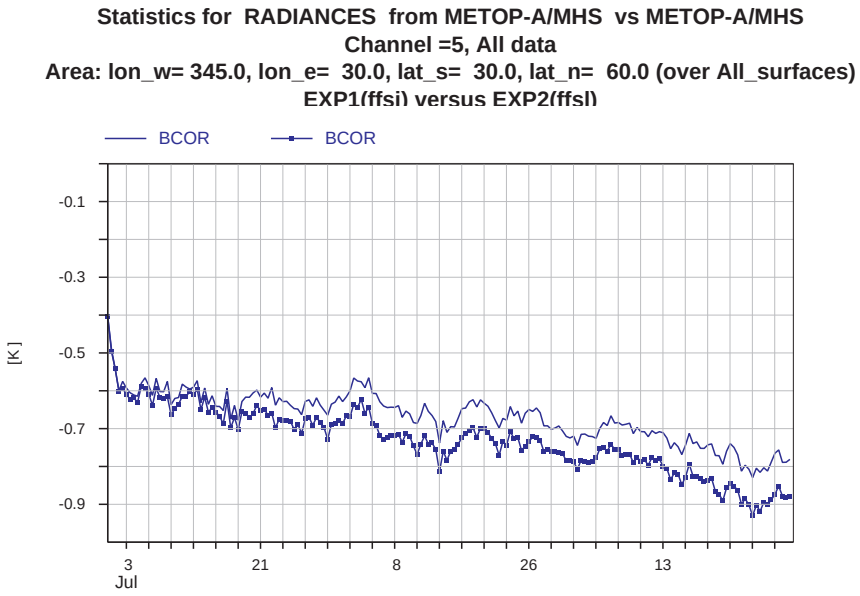


Figure 21: Mean bias correction over the Northern hemisphere for the experiments with 5% (purple dashed line) and 100% GPSRO data usage (green dashed line).The panels show the results for METOP AMSU-A (top left), METOP MHS (bottom left), METOP HIRS (top right) and AQUA AMSR-E (bottom right) for each assimilated channel. Statistics computed from 00 and 12UTC analyses between 7 July and 30 September 2008.

5.1.2 Synergy between radiosondes and GPSRO as temperature anchors

As it has been mentioned for the radiosonde impact investigation with 66% GPSRO data coverage, the presence or absence of radiosondes did not show much impact on the bias correction statistics for any satellite data. When the GPSRO data usage is reduced some sensitivity of the bias correction to radiosonde usage can be observed, but it is still much weaker than the sensitivty of bias correction to GPSRO usage itself. The latter is illustrated in Figure 21, where the purple and green dashed lines show the mean bias correction for the experiments with 5% and 100% GPSRO data usage, respectively. The 4 panels show the results for METOP AMSU-A (top left), METOP MHS (bottom left), METOP HIRS (top right) and AQUA AMSR-E (bottom right). For these instruments the mean bias corrections that are computed dynamically by the variational bias correction show detectable differences, much larger than the differences due to radiosonde usage even in the case when no GPSRO data are present as anchor. It is interesting to look at how the dynamically estimated biases for different satellite instruments/channels evolve in time in the different denial experiments. Figure 22 shows the 3 month time evolution of VarBC bias correction spatially averaged over Europe for METOP MHS channel-5, which is sensitive to mid-tropospheric humidity and AMSU-A channel-13, which is a high stratospheric temperature sensitive channel. The two curves belong to the expriments with 5% and 100% GPSRO usage, respectively. It can be seen that the MHS bias correction is much less sensitive to GPSRO usage than the higher stratospheric AMSU-A channel bias correction. For the latter the bias correction difference grows fast in the beginning, only after a few cycles it is as large as the bias correction itself and it further grows slowly afterwards during the full 3 month period. For MHS several weeks of assimilation are needed to detect any difference and the differences remain moderate during the whole period. Tropospheric peaking AMSU-A channels are clearly less sensitive to GPSRO usage than stratospheric ones. The change of bias correction of temperature sensitive satellite measurments due to radiosonde usage seems to be smaller than due to GPSRO and for humidity sensitive MHS data this change is also moderate. Figure 23 shows the time evolution of spatially averaged VarBC bias correction for MHS channel-4 and for AMSU-A channel-6 (a mid-tropospheric temperature channel) for 3 experiment pairs: 5% vs 100% GPSRO; radiosonde denial vs control with 5% GPSRO usage; radiosonde denial vs control with 66% GPSRO usage. Here the time period covers 6 weeks after a one week warm-up of each experiment. It is seen that for MHS the sensitivity on radiosonde usage is only marginally larger than on GPSRO usage. On the other hand the GPSRO sensitivty is larger than the radiosonde sensitivity for AMSU-A, even for the tropospheric channel-6. The radiosonde sensitivity is, on the other hand slightly increasing when less GPSRO data are present to anchor the biases.



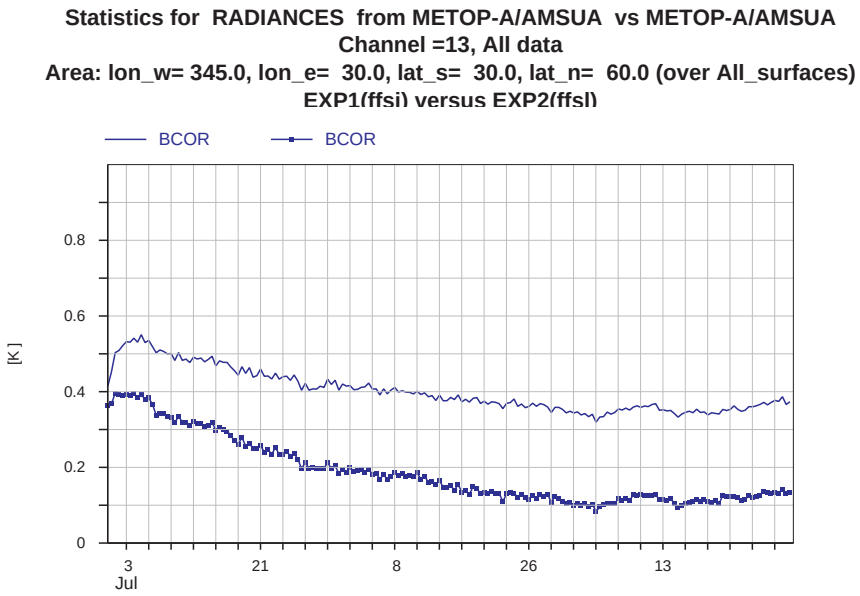


Figure 22: 3 month time evolution of VarBC bias correction spatially averaged over Europe for METOP MHS channel-5 (top) and AMSU-A channel-13 (bottom). The two curves belong to the expriments with 5% and 100% GPSRO usage, respectively.

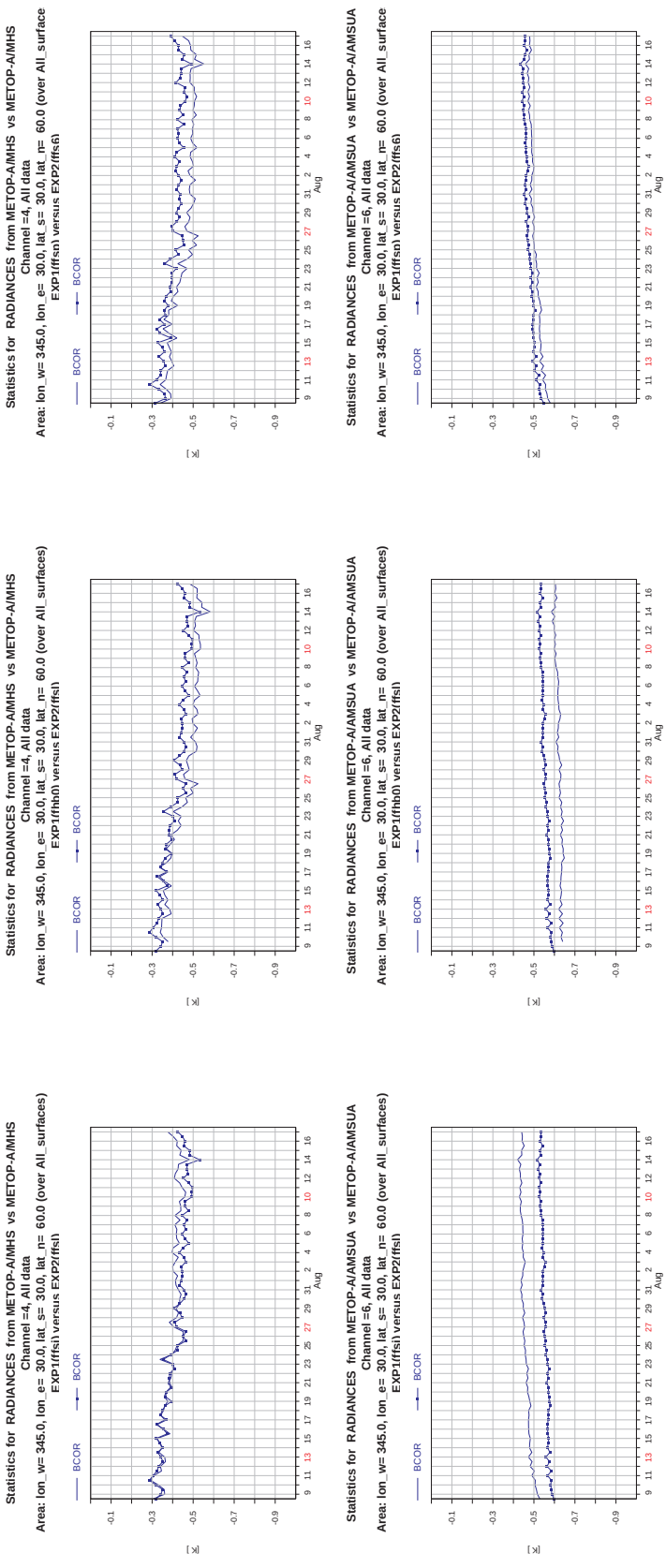


Figure 23: Time evolution of VarBC bias correction spatially averaged over Europe for METOP MHS channel-4 (top) and for AMSU-A channel-6 (bottom), for 3 experiment pairs: 5% vs 100% GPSRO (left); radiosonde denial vs control with 5% GPSRO usage (middle); radiosonde denial vs control with 66% GPSRO usage (right). Time period covers 6 weeks after a one week warm-up of each experiment.

5.1.3 *Forecast impact of radiosonde, aircraft and GPSRO data denial experiments against different observations: Forecast departure statistics*

Assessing the impact of different observation types on the forecast performance has so far been presented by standard verification tools that are usually applied at ECMWF. These compare forecasts with verifying analyses (the operational analysis in this case). As mentioned earlier, verification against operational analysis, own analysis or against observations can produce rather different results and can lead to opposite conclusions as it has been found on a number of occasions. A typical example is when a new observation type is introduced to the assimilation system, that may add systematically larger increments of smaller scale to the analysis, making verification against analyses questionable. Verifying the forecasts against observations will provide a more objective veirfication in such cases. The standard verification package used at ECMWF contains this option, but it is limited to the use of radiosonde observations. Therefore conclusions from this type of verification can not be drawn for areas that are poorly covered by radiosondes (e.g. oceans).

A very powerful tool to assess the performance of the model, especially the performance of the data assimilation system, is to compare both the analysis and the first-guess fields with all available observations (conventional and satellite based). This comparison is perfomed in observation space and uses the same observation operators (temporal integration, interpolation scheme, radiative transfer model) as the data assimilation scheme itself. The results of this comparison are saved as first guess and analysis departures and the, so called, OBSTAT package is used to compute statistics and visualize them in different forms.

A new configuration that has been recently developed is the generalization of the first guess and analysis departure computations so that they can be performed against forecasts of any time range. Some results obtained by this new method for the radiosnde, aircraft and GPSRO denial experiments are presented below.

In this study the forecast departures have been derived for forecast time ranges of 48 and 96 hours. Figure 24 shows the evolution of the standard deviation of forecast departures over the Northern hemisphere for different observation types: radiosonde temperatures (upper left), aircraft temperatures (bottom left), METOP AMSU-A radiances (top right) and METOP MHS radiances (bottom right). Each panel contains departure standard deviation curves that represent analysis (red dotted), first guess (red continuous), 48-hour forecast (black continuous) and 96-hour forecast departures (black dotted).

The departure statistics clearly show the error growth with increasing forecast lead time both against conventional temperature observations and against temperature and moisture sensitive satellite observations. The vertical structure of the temperature errors shows a sharp error maximum near the tropopause level, especially towards longer lead times. The radiosonde temperature statistics are comparable with the standard verification scores against radiosondes (not shown here). However it has to be noted that even for radiosonde temperatures the forecast departure configuration is expected to give more accurate results for a number of reasons: (1) being embedded in the data assimilation system it uses the same quality control criteria as the 4D-Var system itself; (2) radiosonde measurements are bias corrected; (3) more data in the vertical, including significant level observations, are used to derive the statistics.

Figure 25 shows the impact of radiosonde data denial on forecast departure statistics for the Northern hemisphere and for the same observations as Figure 24. The curves represent the difference of error standard deviations between the denial and the reference experiment (again with 66% GPSRO data usage). Positive values indicate a positive impact of radiosonde data.

Against satellite data the first guess and analysis departure statistics are also shown in addition to the 48 and 96-hour forecast departure standard deviations. Both the radiosonde and aircraft temperature depar-

ture statistics confirm that the impact of radiosondes is positive over the examined forecast range and the positive impact is growing with lead time in the upper troposphere (between 200 and 700 hPa). The statistics show a cca. 0.02-0.03K error reduction against radiosonde and aircraft data due to radiosonde usage in the lower part of troposphere, while it is much larger near 200hPa (0.08K). The statistics against radiosonde temperatures shows that in the stratosphere radiosonde assimilation has a smaller but still positive impact, however this positive impact is detectable only up to 2 days of lead time. Radiosonde wind statistics over the Northern hemisphere show a similar structure. One exception is stratospheric wind statistics over the Tropics, that deserves special attention (discussed in the next paragraph). Radiosonde data assimilation is reducing the forecast error against AMSU-A and MHS data as well. Here, the impact on first-guess and analysis departures is negative for certain channels (e.g. MHS channel 4 and AMSU-A channel-8), i.e. the presence of radiosonde data may reduce the ability of the assimilation to fit the satellite data, but the forecast statistics clearly indicate their positive impact. For AMSU-A the impact is growing with lead time for every channel, while for the humidity sensitive MHS the 48-hour is larger than the 96-hour impact, which confirms that humidity forecast impacts are restricted to shorter time scales. The error reduction with respect to satellite data is 0.01-0.03K.

It has been mentioned above that the influence of radiosonde observations in the stratosphere is somewhat different from the tropospheric impacts. It is worth to look into this question by comparing the forecast departure statistics of the radiosonde denial, the stratospheric radiosonde denial (i.e. denial of radiosonde data above 50hPa) and the control experiments. Figure 26 shows the impact of complete radiosonde data denial (continuous lines) and the impact of stratospheric radiosonde denial (dashed lines) on forecast departure statistics for the Northern hemisphere, for 48h forecast (red curves) and for 96h forecast (green curves). Verification is against radiosonde temeratures on the left panel and against METOP AMSU-A brightness temperatures on the right panel. As expected, when only stratospheric radiosonde data are denied, no impact on scores against radiosonde temperatures can be seen in the troposphere. That is not the case higher in the atmosphere, between 100-20hPa, where stratospheric radiosonde data seem to have a positive forecast impact. When verification is done against METOP AMSU-A data (right panel) it can be seen that the stratospheric denial does not influence the forecast fit to tropospheric and lower stratospheric peaking channels, i.e. the channels for which the total radiosonde denial shows the largest impact. It is only channel-12 and 13, where stratospheric radiosonde data assimilation seems to have detectable positive impact on forecast fit to AMSU-A. As mentioned before, the situation is different for wind forecasts. So far statistics have been shown for the Northern hemisphere, where the highest density of radiosonde observations is present and therefore where the impact is the largest. This is not true for wind forecast impact, especially in the stratosphere. Figure 27 shows the standard deviation and bias of zonal wind component fit to radiosonde winds for forecast ranges of 48h (continuous) and 96h (dotted) for the control experiment (red) and for the stratospheric radiosonde denial experiment (black) over the Tropics. Here the experiment pair with 5% GPSRO usage has been chosen because in this case the signal is slightly stronger than with the denser GPSRO coverage. The impact of stratospheric radiosonde observations is clearly positive and they highly reduce both the random and the systematic component of the forecast error in the stratosphere. The positive impact is stronger after 48 hours , but it is still significant after 4 days. Verification of stratospheric vector wind against operational analysis in the Tropics shows the same signal (Figure 28). In the extra-tropics the wind impact is moderate (not shown here) and it is similar to the temperature signal presented before.

Figure 29 shows the same statistics as Figure 25 but for the aircraft denial experiment. Similarily to the radiosonde data, the aircraft data also prove to have positive forecast impact with respect to all shown four observation types. There are, however, some differences:

- near 200 hPa the temperature error growth is faster due to the large amount of denied flight level

data;

- radiosondes have a larger impact than aircraft data on the 48-hour forecast errors in the lower troposphere, while in the higher troposphere the opposite can be observed;
- on the 96-hour forecast range radiosondes seem to have larger impact everywhere and this can be observed not only against conventional temperature observations but against METOP AMSU-A data as well;
- in terms of first-guess and forecast errors as compared to METOP MHS data, the aircraft data denial seems to have the larger impact. This is rather surprising given that humidity is measured by radiosondes and not measured by aircrafts (the same surprising result has been mentioned related to Figure 6).

Since satelllite measurements are available at high spatial density it is possible to examine the geographical distribution of the forecast impact against satellite radiances, i.e. the spatial distribution of the data used for the curves in the previous figures. Figure 30 shows this geographical distribution over a domain covering most of Europe and North-America for the radiosonde impact experiment and the 96-hour forecasts as measured against METOP AMSU-A channel-6 (top) and METOP MHS channel-3 (bottom). Positive values (brown to red colors) indicate a positive forecast impact of the radiosonde assimilation against satellite data. For both verifying satellite measurements the impact of radiosondes is clearly positive, especially over Europe, the Eastern coast of America and some parts of the Atlantic.

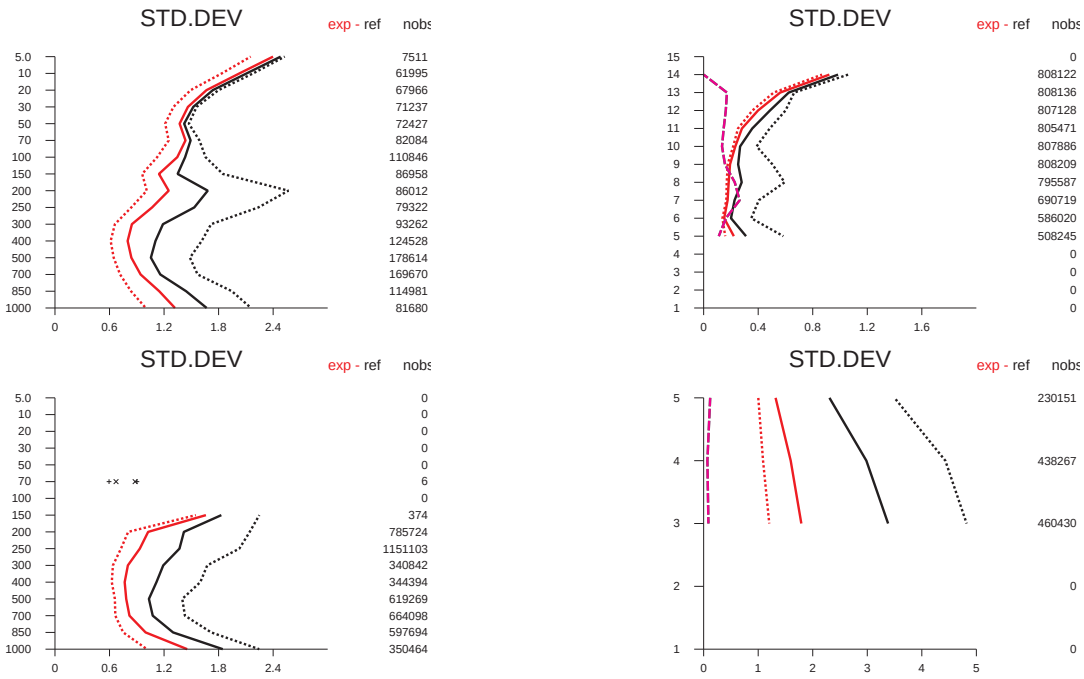


Figure 24: Evolution of the standard deviation of forecast departure over the Northern hemisphere for different observation types: radiosonde temperatures (upper left), aircraft temperatures (bottom left), METOP AMSU-A radiances (top right) and METOP MHS radiances (bottom right). Each panel contains 4 departure standard deviation curves that represent analysis departure (red dotted), first guess departure (red continuous), 48 hour forecast departure (black continuous) and 96 hour forecast departure (black dotted) statistics. Results have been derived from the reference experiment with 66% GPSRO usage. Statistics computed from 00 and 12UTC forecasts between 10,July-17,August , 2008.

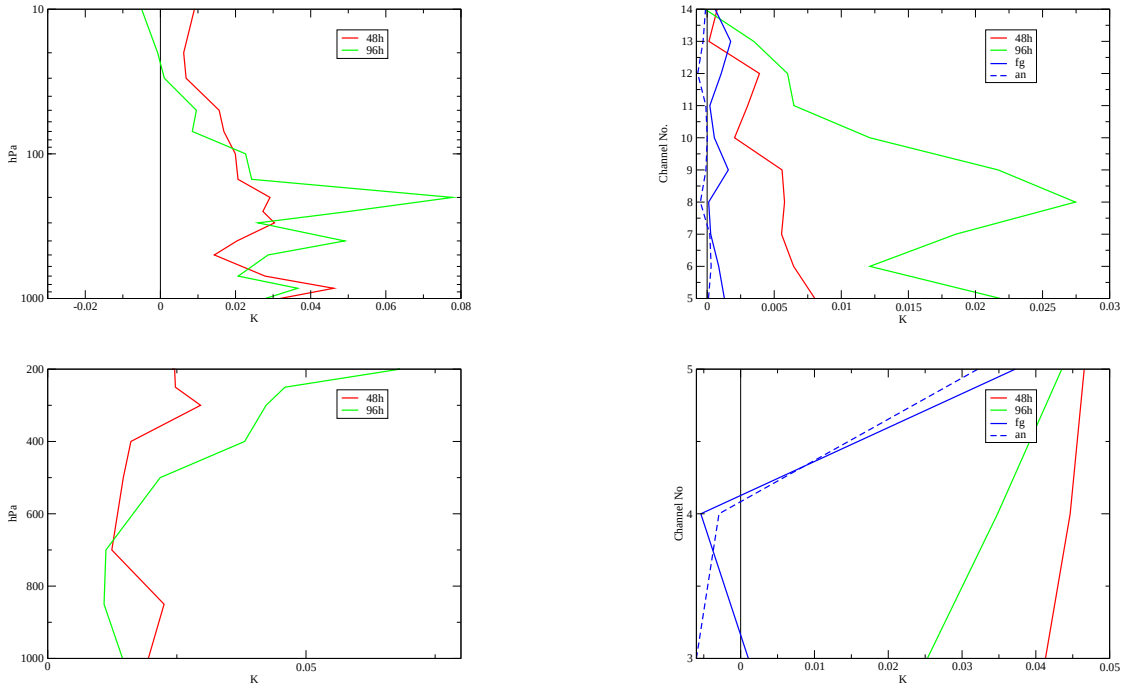


Figure 25: Evolution of the difference of standard deviation of forecast departure over the Northern hemisphere for different observation types: radiosonde temperatures (upper left), aircraft temperatures (bottom left), METOP AMSU-A radiances (top right) and METOP MHS radiances (bottom right). Left panels contain 2 and right panels 4 departure standard deviation curves that represent analysis departure (blue dashed), first guess departure (blue continuous), 48 hour forecast departure (red continuous) and 96 hour forecast departure (green continuous) statistics. Results have been derived from the radiosonde denial and the reference experiment with 66% GPSRO usage. Statistics computed from 00 and 12UTC forecasts between 10,July-17,August , 2008.

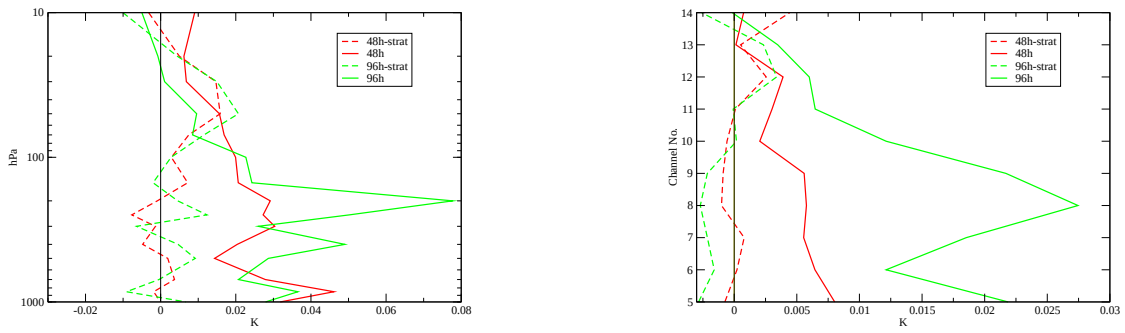


Figure 26: Evolution of the difference of standard deviation of forecast departure over the Northern hemisphere for different observation types: radiosonde temperatures (left) and METOP AMSU-A radiances (right). Both panels contain 4 departure standard deviation curves that represent difference between radiosonde denial and control (continuous lines) and between stratospheric radiosonde denial and control (dashed lines) experiments for the 48h forecast range (red curves) and for the 96h forecast range (green lines). The experiments used 66% GPSRO coverage. Statistics computed from 00 and 12UTC forecasts between 10,July-17,August , 2008.

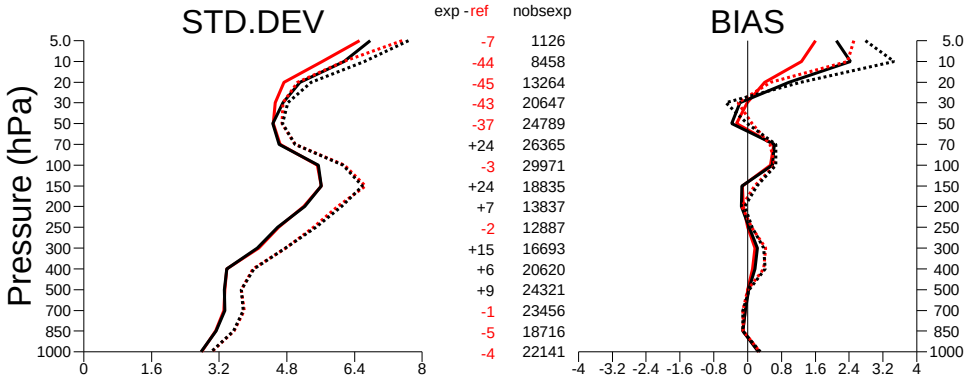


Figure 27: Fit of 48h forecast (solid) and 96h forecaast (dashed) to radiosonde zonal winds in the Tropics from the stratospheric radiosonde denial (black) and Reference (red) experiments with 5% GPSRO data coverage. Left panel shows standard deviations, right panel shows biases. Statistics computed from 00 and 12UTC forecasts between 7 July and 17, August, 2008.

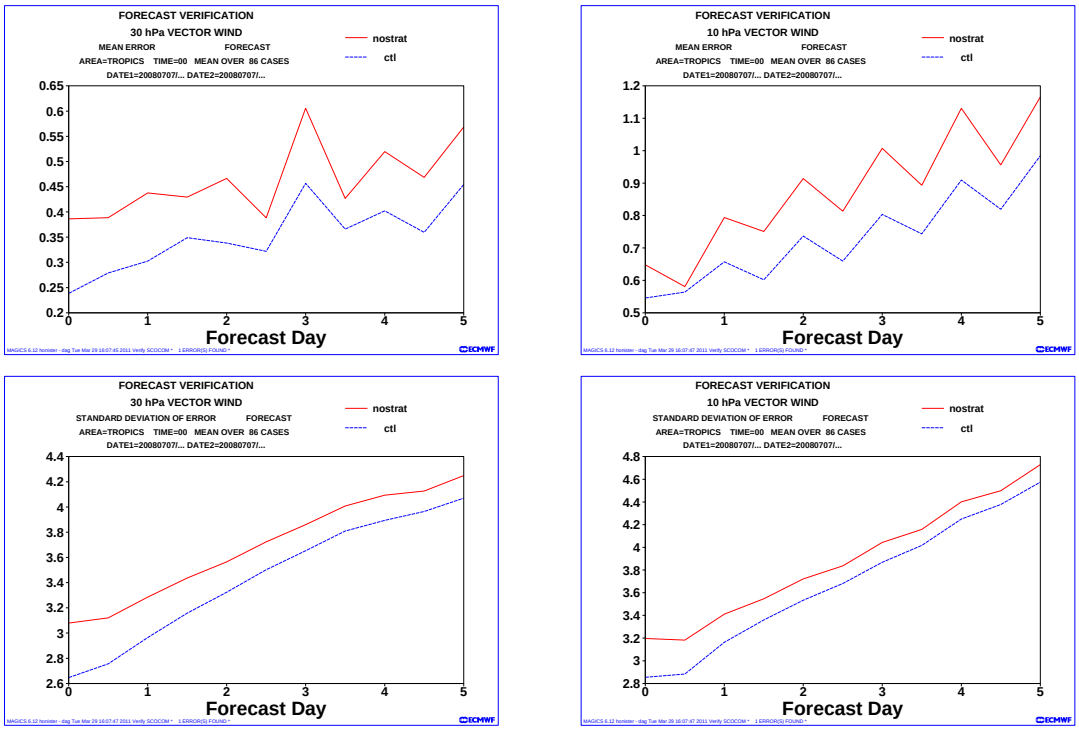


Figure 28: Tropical vector wind forecast biases (top) and error standard deviations (bottom) against operational analysis on two stratospheric levels, i.e. on 30hPa (left) and 10hPa (right). Curves represent two experiments: control with 5% GPSRO usage (blue curves) and stratospheric radiosonde denial with 5% GPSRO usage (red curves). Statistics computed from 00UC forecasts between 7 July and 30, September, 2008.

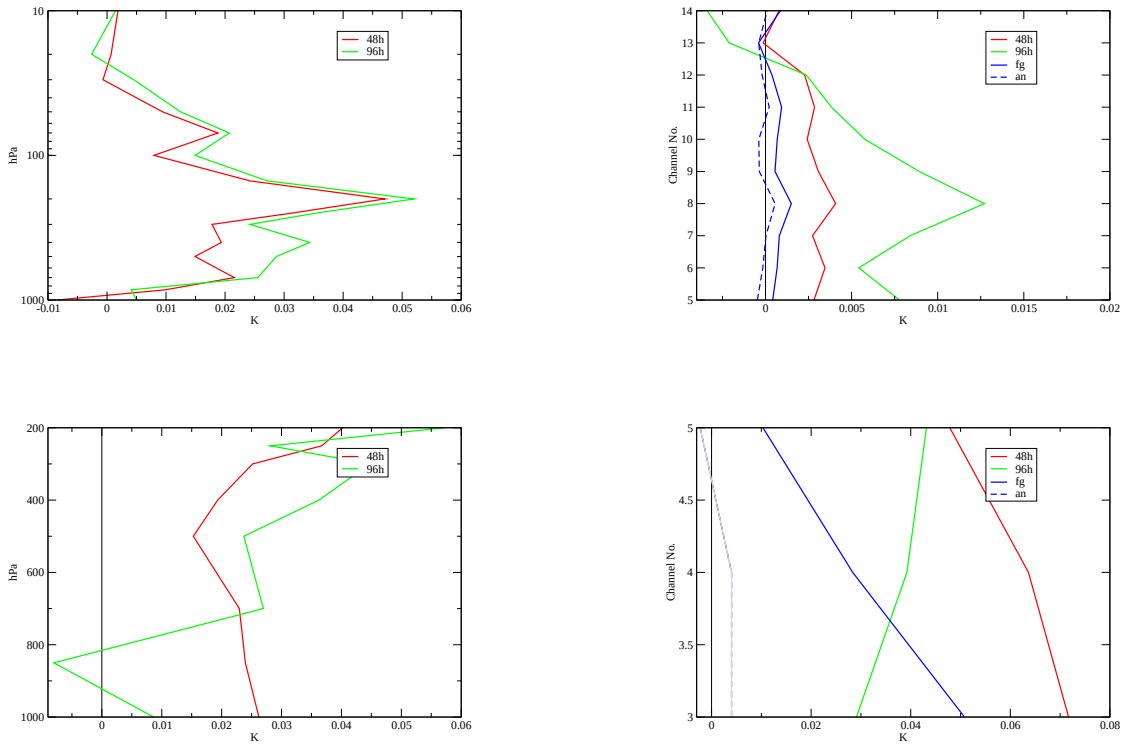


Figure 29: Same as Figure 25 but for aircraft data denial

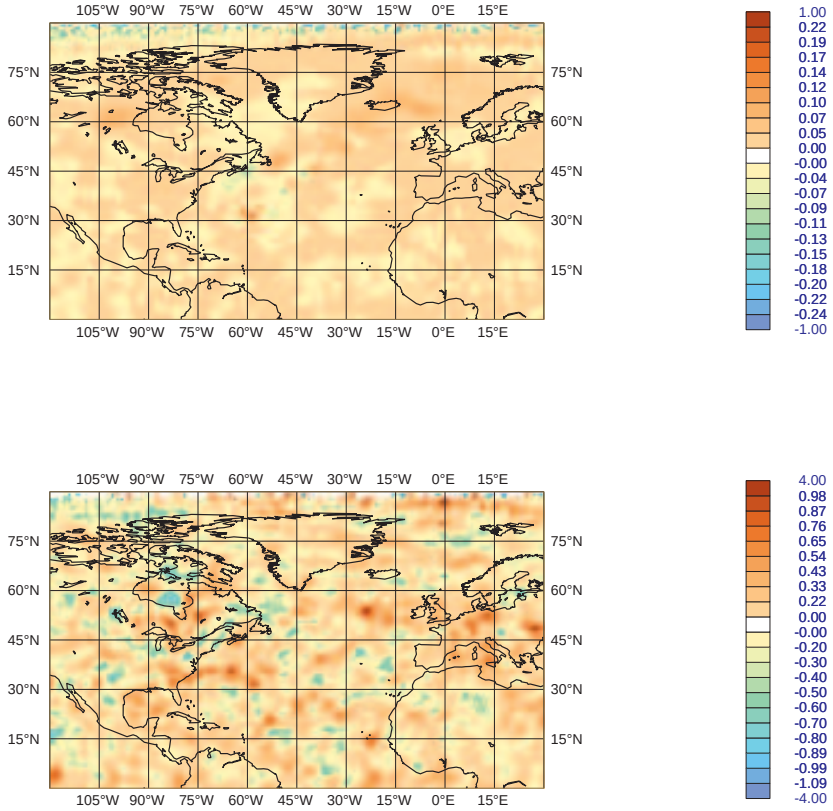


Figure 30: Geographical distribution of radiosonde impact over a domain containing most of Europe and North-America, measured as the difference of standard deviation of 96 hour forecast departure from METOP AMSU-A channel-6 (top) and METOP MHS channel-3 (bottom) observations. Results have been derived from the radiosonde denial and the reference experiment with 66% GPSRO usage. Statistics computed from 00 and 12UTC forecasts between 10,July-17,August , 2008. Positive values (brown to red colors) indicate a positive forecast impact of radiosonde assimilation against satellite data.

5.1.4 Summary

A set of OSEs has been performed to investigate the interaction of radiosonde, aircraft and GPSRO measurements in terms of their analysis and forecast influence. The synergy of these observations as anchors for the bias correction of other observations was also studied. A newly developed tool to verify forecasts against conventional and satellite observations was applied and it proved to give important additional information on the observation impacts. The set of OSEs consisted of radiosnde and aircraft denial experiments performed with successively reduced GPSRO density. Stratospheric radiosonde impacts were additionally studied with denial of radiosonde data above 50hPa.

The aircraft denial experiments showed a remarkable influence on the mean analysis state. As expected, this is the strongest near the cruise level (around 250hPa) over the Northern hemisphere, where the change of mean temperature analysis reaches 0.4K. This is clearly related to a warm temperature bias of aircraft measurements that can be also detected in the analysis and first guess fit to GPSRO bending angles. In spite of the strong temperature biases, aircraft data prove to have a significantly positive impact on forecast scores.

Radiosonde data also have an imapct on the mean analysis state, mostly over large continental areas. The temperature impact is a warming in the boundary layer and a cooling effect above. Near the tropopause the impact is small in spite of the fact that mean radiosonde temperature departures are the largest there. This is likely due to the overwhelming dominance of aircraft data near the cruise level. Radiosonde assimilation has a clearly visible impact on the mean humidity state too, that is an overall moistening of the analysis over most of the continents. This is the integral result of a dominant moistening in the lower troposphere and a drying impact above. Radiosonde data, just like aircraft measurements, have a significant positive impact on forecast scores. Here the positive impact is observed for humidity scores as well.

The comparison of aircraft and radiosonde impact on forecast scores suggests that in the higher troposphere aircraft data contribute more to the forecast score improvement and this dominance spreads down to 400-500hPa. Below this, radiosondes are more beneficial on a hemispheric average. Nevertheless, over Europe and North-America, where the largest observation density of both radiosonde and aircraft data profiles are available, the aircraft contribution to the forecast quality is dominant even in the lower troposphere. Stratospheric radiosonde data prove to have a particularily strong positive impact on wind forecast scores, mainly over the Tropics. This was proven both against analysis and against radiosonde data.

As it has been shown in earlier studies, GPSRO data also strongly constrain temperature profiles and surface pressure of the analysis. Therefore it is interesting to see the impact of radiosonde and aircraft data in conjunction with GPSRO data coverage. The OSEs with different GPSRO data densities showed that in the stratosphere GPSRO data have a warming effect while in the troposphere they cool the analysis. This impact is in agreement with radiosonde temperature profiles except for the lower boundary layer. Comparison of GPSRO impacts in conjunction with aircraft and radiosonde denials suggests the impact of GPSRO data on the mean analysis state is mainly driven by aircraft biases.

The synergy between radiosondes and GPSRO as temperature anchors has been examined through the denial experiments. It has been found that the presence or absence of radiosondes did not show strong impact on the bias correction, at least much less than GPSRO data. The small, but still detectable anchoring effect of radiosondes proved to be stronger when less GPSRO data were assimilated.

5.2 Follow-on investigations of the Upper air redesign study

In 2009 a study was prepared to provide input for the definition of a European-wide network of ground-based upper-air observing systems. It was a joint study by ECMWF as the Center running the global model experiments of the different ground-based upper-air observing system scenarios and by three different National NWP Centres that were the partners responsible for performing the corrsponding Limited Area Model experiments using the Lateral Boundary Conditions provided by ECMWF. The main conclusion of the global experiments performed at ECMWF was that the reduction of European radiosonde network to a 100km horizontal spacing keeping simultaneously the full aircraft data usage (scenario 3a-b) does not lead to significant degradation of the global model performance over the European domain. Using a 250km joint thinning distance for radiosondes and aircraft data over Europe (scenario 4) resulted in some slightly controversial conclusions in the two examined seasons and between the different models. Therefore it has been decided that scenario-3b and scenario-4 are revisited in this study using the most up-to-date model cycle and a 6 month continuous period of experimentation (the same model version and an extended period of the other OSEs of this study).

The same evaluation has been performed for the full 6 months as in the 2009 study. Standard verification scores have been computed both against own and against operational analysis. Verification against observations has been performed by computing forecast departure statistics as in subsection 5.1.3. For most of the parameters and vertical levels scenario-3b did not show significant difference with respect to the control experiment in agreement with the findings of the 2009 study. The only parameter where both verification against operational and against own analysis showed a significant deterioration of the forecast due to data denial is 100hPa wind in the first 48h. This is demonstrated on Figure 31 together with scenario-4 (verification done against own analysis). It can be seen that the forecast impact in both scenarios is a deterioration due to data denial. This impact is significant on 95% confidence level up to 1.5 days for scenario-3b and 3 days for scenario-4. For scenario-4 the impact is significant for wind and temperature forecasts throughout the Troposphere and for low level humidity. The significant impact lasts 2-3 days as seen on Figure 32 for 850hPa relative humidity and 500hPa geopotential height against own analysis.

Forecast departure statistics have been computed for a one month period. Figure 33 shows the difference between the error standard deviation of the denial and control experiment, where the error is defined as the forecast departure from radiosonde temperatures (left) and from METOP AMSU-A radiances (right). Positive values indicate that the additional observations reduce the departure standard deviation. The two examined forecast time ranges are 48h (red lines) and 96h (green lines). When compared against radiosonde temperatures, 48h forecasts do not show any clear impact against these observations while at the 96h forecast range degradation due to data denial can be seen. Against AMSU-A radiances the impact can be clearly identified only for scenario-4 at the 96h time-range for channel-8-10, i.e. for the channels sensitive to higher tropospheric temperatures. If we compare these results with the statistics derived in the same manner for the global radiosonde or aircraft denial (see in subsection 5.1.3) not surprisingly the impacts are much smaller here.

The impact of radiosonde and aircraft measurement density on the performance of precipitation forecast has also been quantified for the entire 6 months of experimentation. Precipitation forecasts were compared to rain gauge measurements over Western Europe (domain corners: SW:-10W,40N NE:25E,70N). Three different scores were computed, the standard Equitable Threat Score (ETS) and Peirce Skill Score, (hereafter PSS, see e.g. Wilks, 1995) and the Stable Equitable Error in Probability Space (hereafter SEEPS, see Rodwell et al., 2010), that was recently developed and introduced at ECMWF for monitoring precipitation forecasts. The latter has the advantage that it takes into account the climatologial

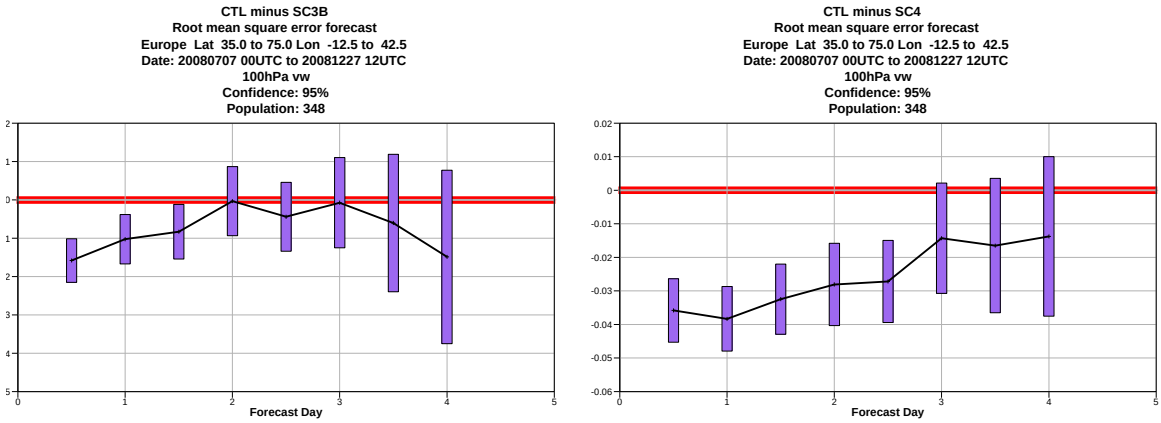


Figure 31: RMS error differences between control and Sc3b scenario (left) and between control and Sc4 scenario (right) for 20080707-20081227 for 100hPa vctor wind. Vertical bars represent significance interval on 95 percent confidence level. Verification is derived from 00 and 12UTC forecasts and it is performed against own verifying analysis. Verification domain is Europe.

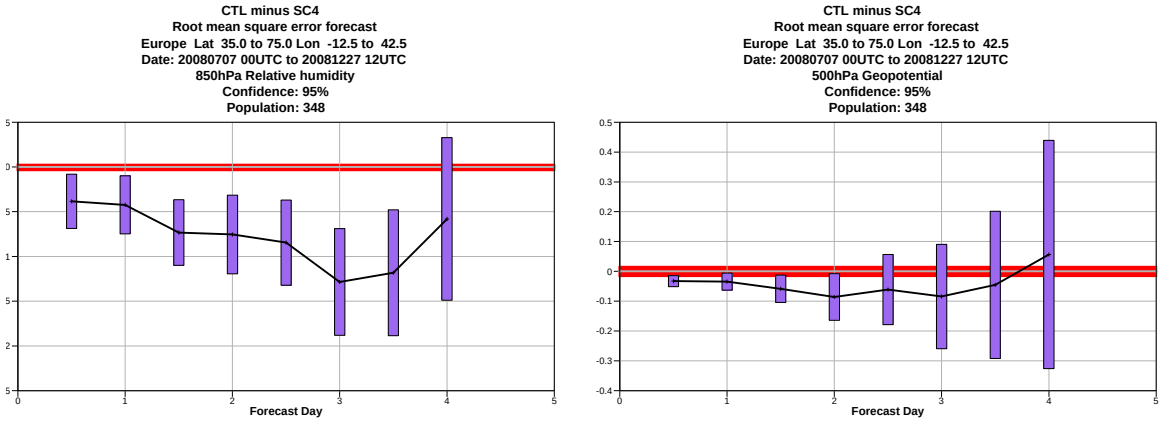


Figure 32: RMS error differences between control and Sc4 scenario for 850hPa relative humidity (left) and for 500hPa geopotential height (right) for 20080707-20081227. Vertical bars represent significance interval on 95 percent confidence level. Verification is derived from 00 and 12UTC forecasts and it is performed against own verifying analysis. Verification domain is Europe.

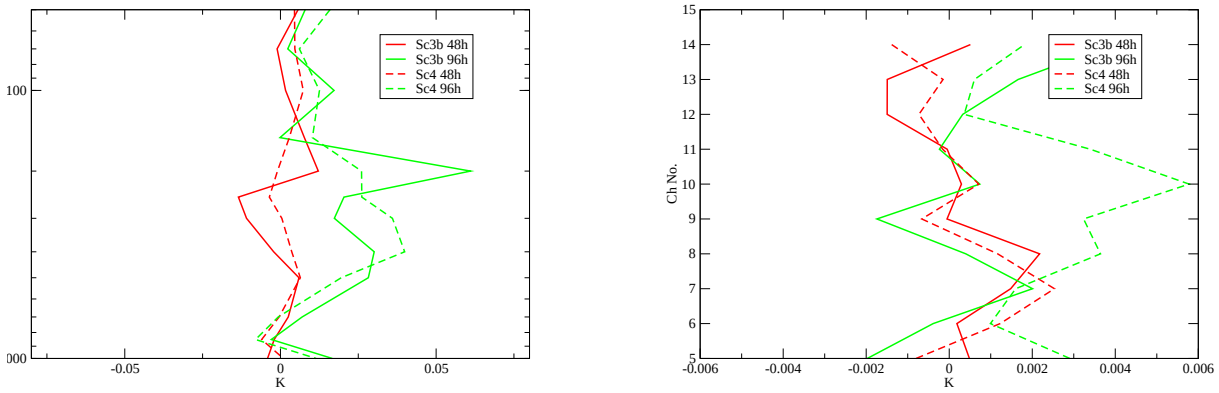


Figure 33: Evolution of the difference of standard deviation of forecast departure over Europe for different observation types: radiosonde temperatures (left) and METOP AMSU-A radiances (right). Both panels contain 4 departure standard deviation curves that represent difference between scenario-3b and control (continuous lines) and between scenario-4 and control (dashed lines) experiments for the 48h forecast range (red curves) and for the 96h forecast range (green lines). Statistics computed from 00 and 12UTC forecasts between 10,July-17,August , 2008.

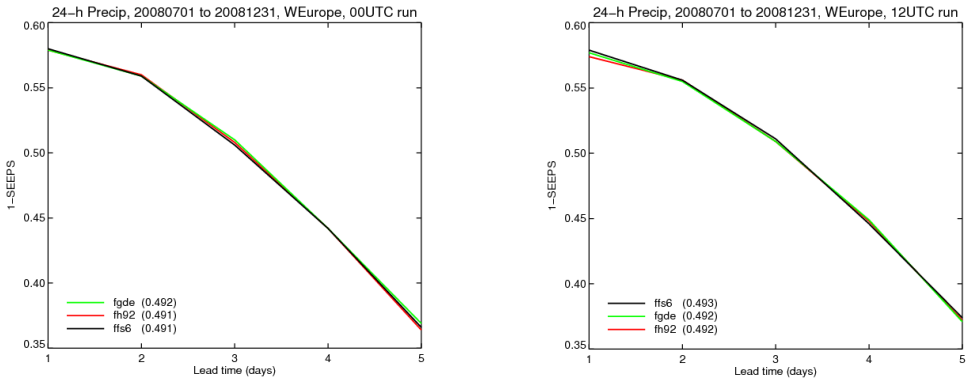


Figure 34: SEEPS for the 00 UTC (left) and the 12 UTC run (right) for Western Europe as a function of lead time for the control (black) and the two data denying experiments (green, red). The experiment names are control: ffs6 (black), scenario 3b: fgde (green) and scenario 4: fh92 (red).

cumulative distribution function of precipitation. The three experiments on the precipitation verification figures are the control (experiment id: ffs6, black curves), scenario 3b (exp id: fgde, green curves) and scenario 4 (exp id: fh92, red curve). From the definition of the three scores it follows that larger values represent better precipitation forecasts. Figure 34 shows the SEEPS scores for the three experiments from forecast lead time of 1 day to 5 days. The 00UTC (left) and 12UTC (right) forecasts are shown separately. Overall, the data denial has little effect on the SEEPS scores and it is within the sampling uncertainty. The only systematic effect is seen for the 12UTC run at day 1. It has to be underlined that at 00UTC the data usage of scenario 3b is identical to the control, therefore it is not surprising that no impact is seen there. However, for scenario 4 the 00 and 12UTC analyses are identical in terms of data selection criteria. The temporal evolution of the day 1 SEEPS scores can be seen on Figure 35. Again for the 12UTC forecasts, the observation surplus tends to have a small positive impact throughout the whole period, while for the 00UTC forecasts the impact is not systematic at all. The ETS (left) and PSS (right) scores are shown for the 12UTC forecasts as a function of forecast lead time on Figure 36 and Figure 37. Figure 36 belongs to 1*mm* precipitation threshold while Figure 37 uses 5*mm* threshold value. Again, little systematic effect is found. The different ranking of the three experiments for thresholds of 1*mm* and 5*mm* suggests that light precipitation is affected differently from stronger precipitation. However, the differences are mostly too small to be significant.

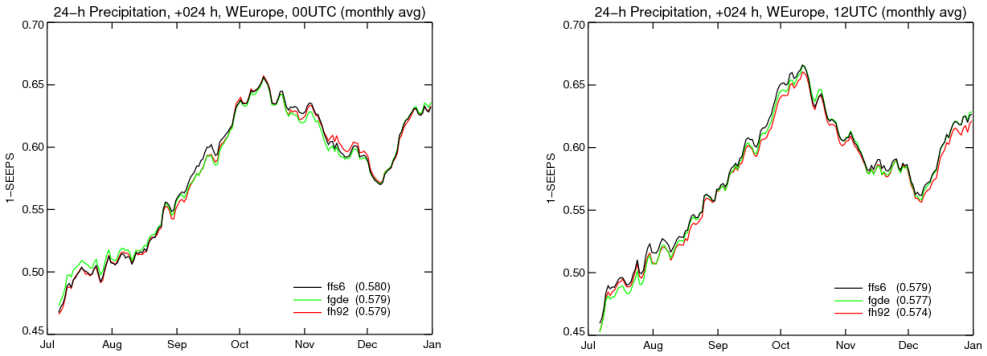


Figure 35: Time evolution of SEEPS for the 00 UTC (left) and the 12 UTC run (right) for Western Europe for lead time of 1 day for the control (black) and the two data denying experiments (green, red). The experiment names are control: ffs6 (black), scenario 3b: fgde (green) and scenario 4: fh92 (red).

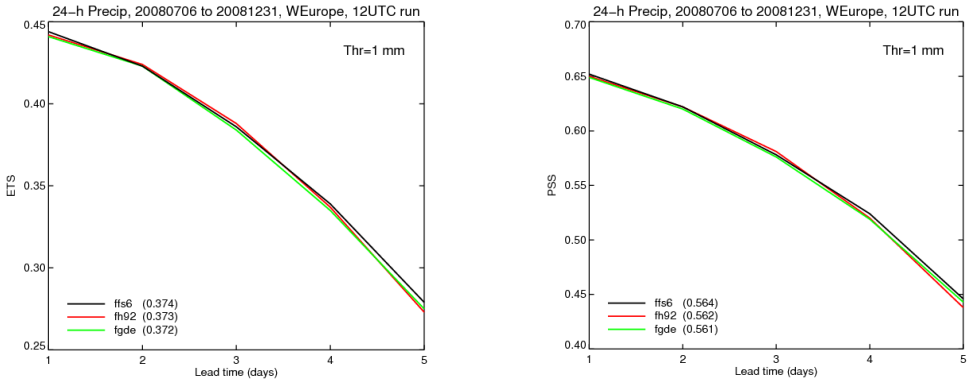


Figure 36: ETS (left) and PSS (right) for a threshold of 1mm for the 12 UTC run for Western Europe as a function of lead time for the control (black) and the two data denying experiments (green, red). The experiment names are control: ffs6 (black), scenario 3b: fgde (green) and scenario 4: fh92 (red).

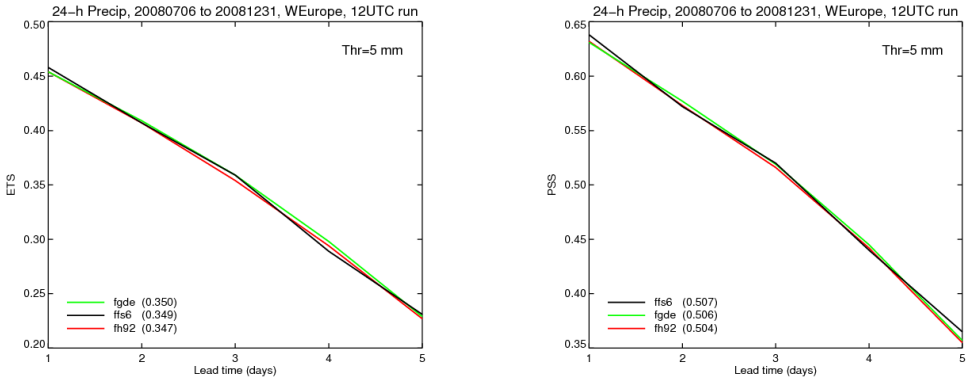


Figure 37: ETS (left) and PSS (right) for a threshold of 5mm for the 12 UTC run for Western Europe as a function of lead time for the control (black) and the two data denying experiments (green, red). The experiment names are control: ffs6 (black), scenario 3b: fgde (green) and scenario 4: fh92 (red).

5.3 Buoy and ship measurement impact studies

Drifting buoy and Voluntary Observing Ship (VOS) measurements represent a very important contribution to the observing system that largely constrain the surface pressure analysis as it will be demonstrated in this work package. They are particularly important because they observe areas where other conventional observations are not available. Recently some new diagnostic tools have been developed at ECMWF that estimate forecast error sensitivity to observations (FSO) and the so called forecast error contribution (FEC) (Cardinali, 2009a,2009b). The FEC values can be grouped by observation types and displayed in form of percentage and then it represents the percentage of total forecast error reduction that is due to the given observation type. Figure 38 shows for 24h forecasts the distribution of FEC in terms of percentage for the different observation types summed up (top) and normalized for a single observation (bottom). This has been produced from our recent operational data assimilation system. The top figure shows that buoy data are one of the most important components of the conventional observing system and according to the bottom figure the buoy impact per single observation is by far the largest among all different kinds of observations.

One of the main objectives of this work package is to estimate the impact of the additional measurements introduced as a result of the E-SURFMAR project on the forecast performance in a state of the art NWP model. In order to estimate the impact of these data on the forecast performance a set of Observation System Experiments has been performed. The OSEs have run in the winter 2008/2009 and a blacklist of 72 North-Atlantic drifting buoys that were active in this period has been provided by P. Blouch (E-SURFMAR Programme manager) to simulate the marine surface pressure measurement coverage that was typical before the E-SURFMAR programme. This blacklist simultaneously denied all non-synoptic time VOS measurements. Figure 39 shows the geographical distribution of buoys over the North-Atlantic area in a randomly selected analysis time for the un-thinned and thinned case.

This OSE set consisted of 8 experiments:

- Control-1: Operational observation coverage.
- Thinned-1: Buoy network reduced by the 72 buoys, only main synoptic time ship measurements, rest of the observing system as in operational model.
- Baseline-1: No buoy measurement, only main synoptic time ship measurements, rest of the observing system as in operational model.
- Baseline-1plus: No conventional surface pressure measurements, rest of the observing system as in operational model.
- Control-2: Control-1 without GPSRO measurements.
- Thinned-2: Thinned-1 without GPSRO measurements.
- Baseline-2: Baseline-1 without GPSRO measurements.
- Baseline-2plus: Baseline-1plus without GPSRO measurements.

This means that the 4 basic OSE scenarios with respect to conventional surface pressure coverage have been performed with and without GPSRO data. This choice has been made because due to the hydrostatic relationship GPSRO data do not only constrain the temperature profile of the analysis, but simultaneously the surface pressure as well (e.g. Radnoti et al, 2010). Thus, presumably the additional buoy surface

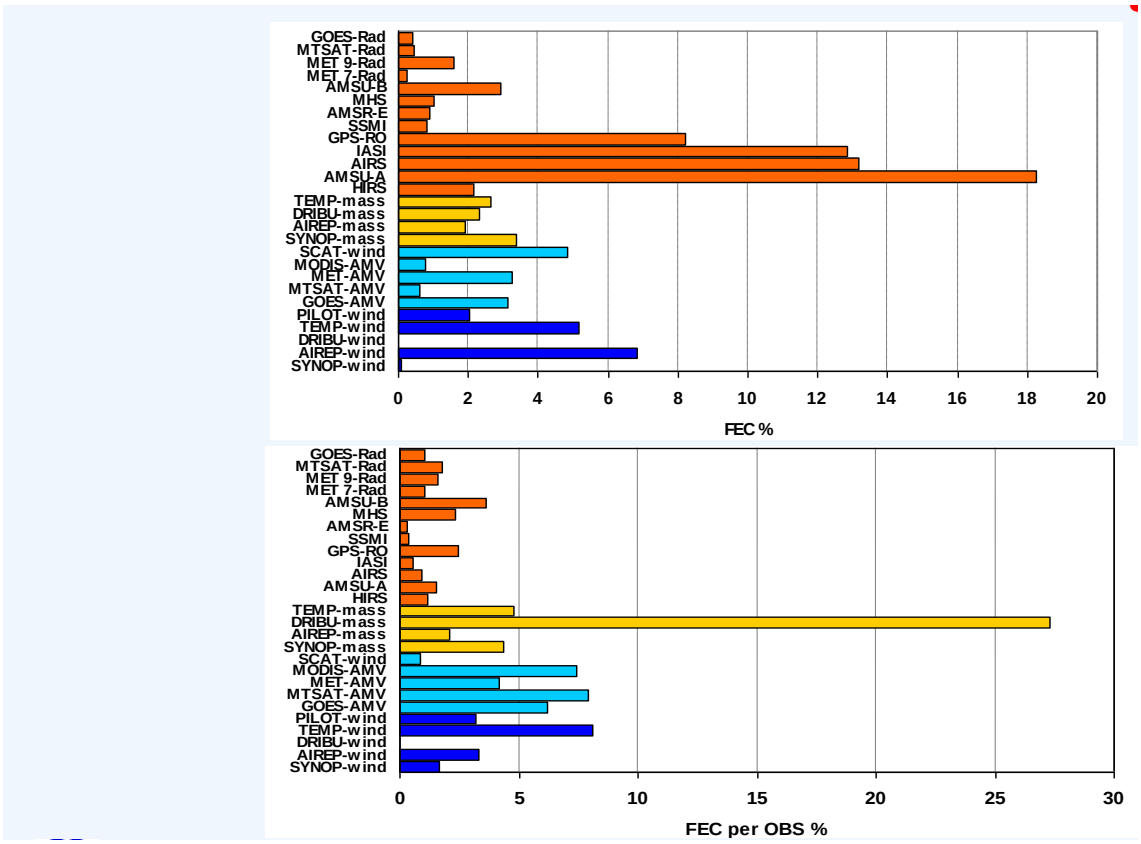


Figure 38: FEC distribution for different observation types summed up (top) and normalized for a single observation (bottom). The FEC contribution is displayed in % and it is derived for 24h forecast errors.

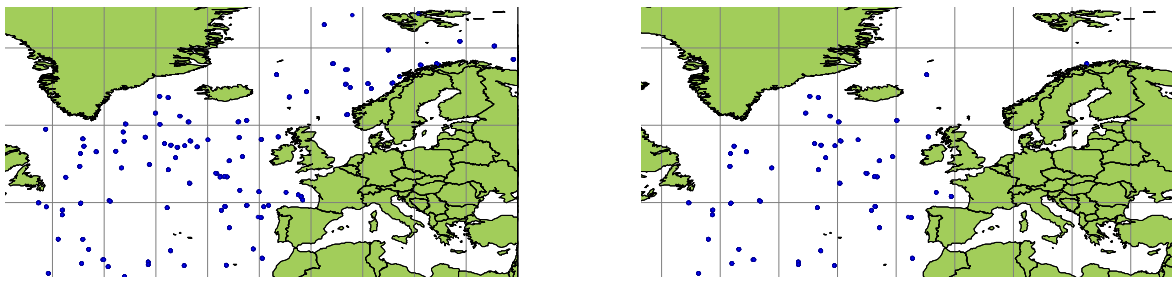


Figure 39: Geographical distribution of assimilated buoys over the North-Atlantic area on 1, January, 2009 within the IFS 4D-Var assimilation system for the control (left) and thinned (right) experiment.

measurements have a more significant impact on model performance when GPSRO data are not present.

These OSEs cover a 2 month period in December, 2008 - January, 2009. The verification has been performed for 7 weeks for each of the experiments ignoring the first week of the period and most of the verifications have been done against the operational high-resolution analyses, that are considered to provide the best available information on the true atmospheric state.

As mentioned earlier, conventional observations, together with GPSRO data play the important role of anchoring the analysis. This is also the case for surface pressure observations where conventional observations from SYNOP, SHIP, METAR and buoy reports and GPSRO data are anchoring the system. This is well demonstrated by Figure 40 which shows the histograms of Southern hemispheric SYNOP (top) and North-Atlantic buoy (bottom) surface pressure analysis and first guess departures. These were derived from a pair of additional experiments where all conventional surface pressure observations were denied and they were used only as passive data for monitoring the assimilation system. The black curves represent the case where GPSRO data have also been denied (scenario Baseline-2plus), while the red curves belong to the case when GPSRO data have been assimilated (scenario Baseline-1plus). The statistics shown on the figure were generated after a 2 months of warm-up assimilation period during which time the assimilation system gradually drifted away from the conventional surface pressure observations. The analysis showed that after this two month period the model stabilized and the statistics shown on Figure 40 became steady. The curves show a significant negative surface pressure model bias (the histogram of observation minus model is asymetric, shifted to the positive direction indicating a negative model bias) both for the analysis and for the first guess when the conventional and GPSRO data are denied. This bias is cca 2.5hPa both against SYNOP surface pressures over land over the Southern hemisphere and against buoy measurements over the Northern Atlantic region when GPSRO data are not assimilated. The assimilation of GPSRO removes half of this bias, but the remaining 1.2hPa systematic error is still significant.

5.3.1 Baseline scenario

To see the basic impact of buoy surface pressure observations on the assimilation and forecast performance first it is worth to compare the baseline experiment with the control both with and without assimilating GPSRO data. Figure 41 shows the histograms of Northern hemispheric (top), Tropical (middle) and Southern hemispheric (bottom) surface pressure analysis and first guess departures from SYNOP/SHIP observations as derived from the Baseline-1 (i.e. buoy denied, GPSRO used, black curves) and Control-1 (i.e. buoy used, GPSRO used, red curves). The somewhat larger sample sizes for the control experiment are due to the fact that in the baseline experiment non-synoptic time ship observations have also been denied in addition to buoy observations. Apart from this, the data usage was identical in the two experiments. It can be seen that due to the use of SYNOP/SHIP data the large biases seen on Figure 40 have disappeared (a cca 0.1hPa negative bias remains over the Northern hemisphere and Tropics, but this is one order of magnitude smaller than the same biases when all conventional surface pressure observations were denied). To see how these residual small biases change with the use of buoy data Table 1 shows the absolute difference of analysis and first guess biases between the baseline (i.e. buoy denial) and control experiments (both with and without GPSRO assimilation) for conventional surface pressure observations (SYNOP and SHIP together and METAR). Units of differences are in *Pa*. Positive values indicate that the bias without assimilating buoy observations is larger. It can be seen that the differences are quite small, but in most of the cases the use of buoy data slightly reduces the biases (with the exception of biases against METAR data when GPSRO is not used). On the Northern

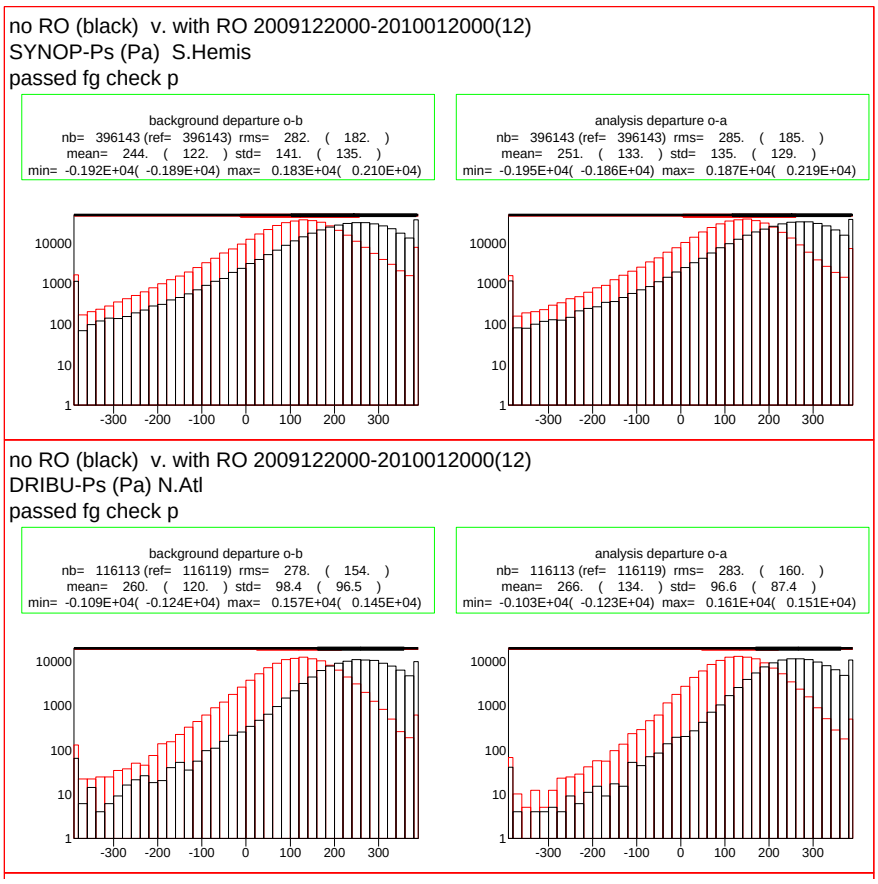


Figure 40: Fit between model first-guess (left) and analysis (right) and independent surface pressure observations in Southern hemisphere SYNOP data (top) and Northern Atlantic buoy observations (bottom) (red curves with and black curves without GPSRO data assimilated). Statistics were generated from period 20/12/2009-20/01/2010 after 2 months of spin-up time to let the model drift away from conventional surface pressure observations.

hemisphere and Tropics this bias reduction is stronger for the short range forecast (first-guess) than for the analysis, which means that the bias reduction impact of the buoy surface observations is partially propagated by the model from the oceans over land. Interestingly, this effect is not found over the Southern hemisphere. Figure 42 shows the geographical distribution of normalized geopotential height RMS error differences between experiments Baseline-2 and Control-2 (i.e. the experiments without GPSRO assimilation) for 1000hPa geopotential height (left) and 700hPa geopotential height (right) for forecast ranges 12h, 24h, 48h and 72h where verification has been performed against operational analysis. The red colors indicate positive impact of buoy data. It can be seen that in the first 24h the positive impact of buoy data is dominant over the sea. Over the Southern hemispheric oceans the RMS error reduction due to buoy data reaches 30%. The positive impact lasts up to 48-72 hours and it is present on 700hPa geopotential as well. Some surprising negative impacts are also seen and it seems to be the strongest at 700hPa geopotential height near the Eastern coast of South-America. Figure 43 shows the same RMS errors without normalization for the 12h time range (1000hPa on the top and 700hPa on the bottom map). The header of the maps shows the mean RMS error reduction over 7 different regions of the globe in $m**2/sec**2$. It can be seen that in certain areas of the Southern oceans the RMS error reduction due to buoy data reaches $50m**2/sec**2$, especially near the sea surface. The areal mean RMS error reduction over the whole Southern hemisphere is $10m**2/sec**2$ on 1000hPa and more than $5m**2/sec**2$ on the 700hPa level, which is remarkable. On other areas the impact is smaller but mostly positive. In the North Atlantic region, on the 1000hPa level the mean RMS reduction is cca $5m**2/sec**2$, but on the 700hPa level we see a small degradation. The earlier mentioned degradation on the 700hPa level at the eastern coasts of South-America reaches $10m**2/sec**2$. A longer experimentation period would be needed to see if these degradations are persistent. Figure 44 shows the zonal cross-sections of normalized root-mean square geopotential height forecast error differences between the Baseline-2 and Control-2 experiments as verified against operational analysis. It can be seen that the positive impact of buoy data is significant up to 2-3 days and the significant positive impact spans through the whole troposphere in the Southern hemisphere while it reaches the mid-troposphere elsewhere (crosses indicate the significance of the results on 95% confidence level). A surprising significant negative impact can be detected over the Tropics, above 600hPa. This negative impact is less expressed in the Baseline-1 versus Control-1 case (not shown), i.e. when GPSRO data are used, but in that case the impact of buoy data is smaller in general. The negative impact is only seen in the RMS error context and it fully disappears for anomaly correlations (not shown). All this suggests that it may be a bias issue and very likely the denial of buoy data results in a cancellation of two opposite biases, i.e. the negative surface pressure bias (seen e.g. on Figure 40) and a positive mid-tropospheric temperature bias (see e.g. in Radnoti et al., 2010). These two have opposite effect on geopotential height and may partly cancel each-other.

As it has been mentioned before, conventional observations and GPSRO data both contribute to the quality of surface pressure analysis. From the two baseline and two control experiments one can estimate how these contributions relate to each-other for different forecast ranges. Figure 45 shows the time evolution of 1000hPa geopotential height RMS error over the Southern hemisphere for these 4 experiments. As expected, the best scores are produced when both GPSRO and buoy data are assimilated (blue curve) and the worst forecast belongs to the poor baseline when both are denied (green curve). The two intermediate cases show an interesting dependence of impact on the forecast range, namely that until day 1 buoy data are more important while after day 2 GPSRO data contribute clearly more to the forecast quality (red and brown curves intersect each-other between time ranges day-1 and day-2). To see the significance of these scores Figure 46 shows for the Southern hemisphere the Control versus Baseline RMS error difference curves with (top) and without (bottom) GPSRO data assimilated. The vertical bars represent the significance interval of the results on 90% confidence level. As it can be seen, the positive impact of buoy data is significant up to 2 days when GPSRO data are present and up to 3 days when they are

Table 1: Absolute difference of analysis and first guess biases between the baseline (i.e. buoy denial) and control experiments (both with and without GPSRO assimilation) for conventional surface pressure observations (SYNOP and SHIP together and METAR). Units of differences are in Pa. Positive values indicate that the bias without assimilating buoy observations is larger.

Obs type, Area	no GPSRO bias-diff analysis	no GPSRO bias-diff first-guess	with GPSRO bias-diff analysis	with GPSRO bias-diff first-guess
SYNOP/SHIP:				
Northern hemisphere	+1.05	+2.68	+1.14	+4.27
Tropics	+0.48	+1.38	+0.77	+2.34
Southern hemisphere	+1.60	-0.4	+2.25	+0.8
METAR:				
Global	-0.1	-1.8	+0.3	+2.7

missing. This is in agreement with all the previous results.

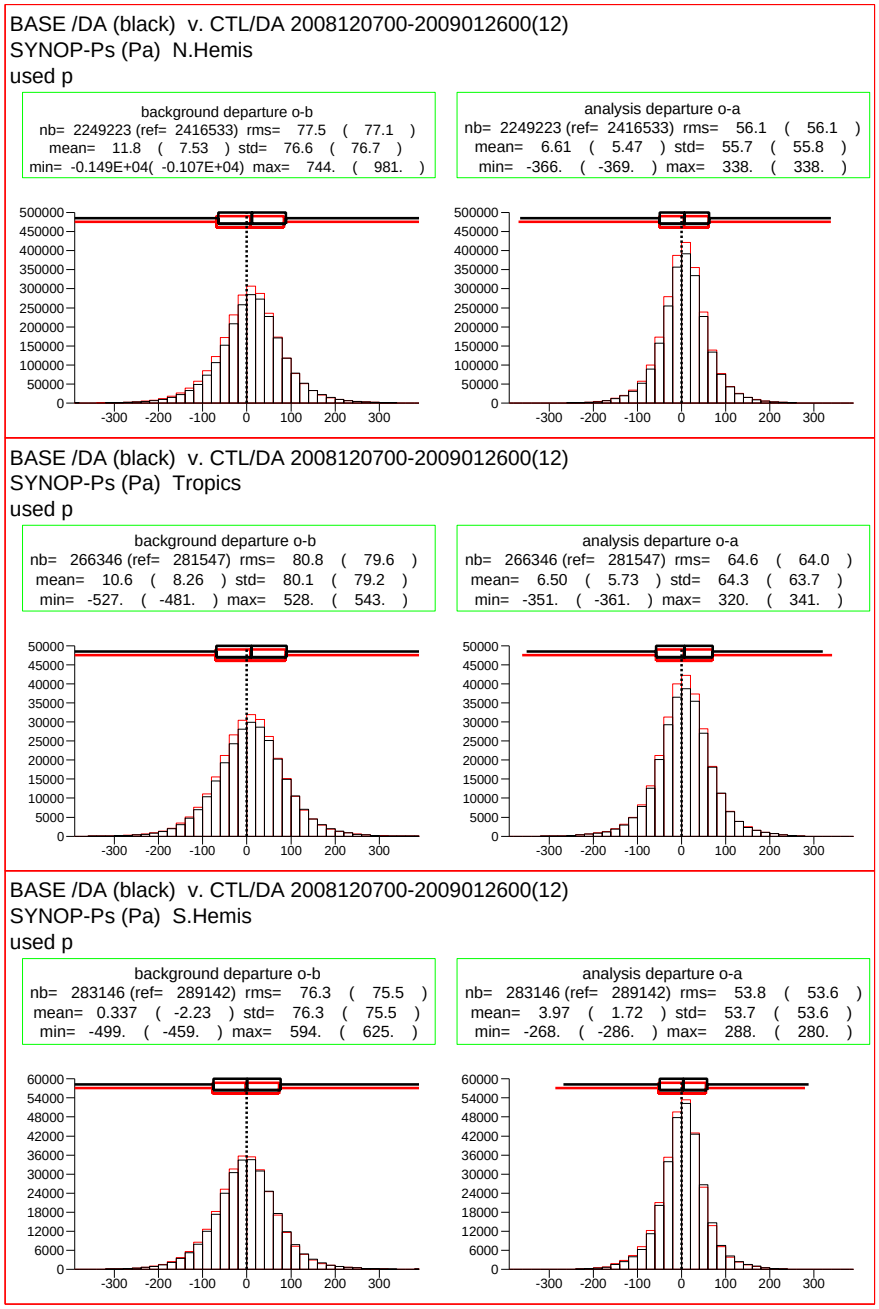


Figure 41: Fit between model first-guess (left) and analysis (right) and SYNOP/SHIP surface pressure observations in Northern hemisphere (top) Tropics (middle) and Southern hemisphere (bottom) for the Baseline-1 experiment (i.e. buoy denied, GPSRO used, black curves) and Control-1 experiment (i.e. buoy used, GPSRO used, red curves). Statistics were generated from period 07/12/2008-26/01/2009.

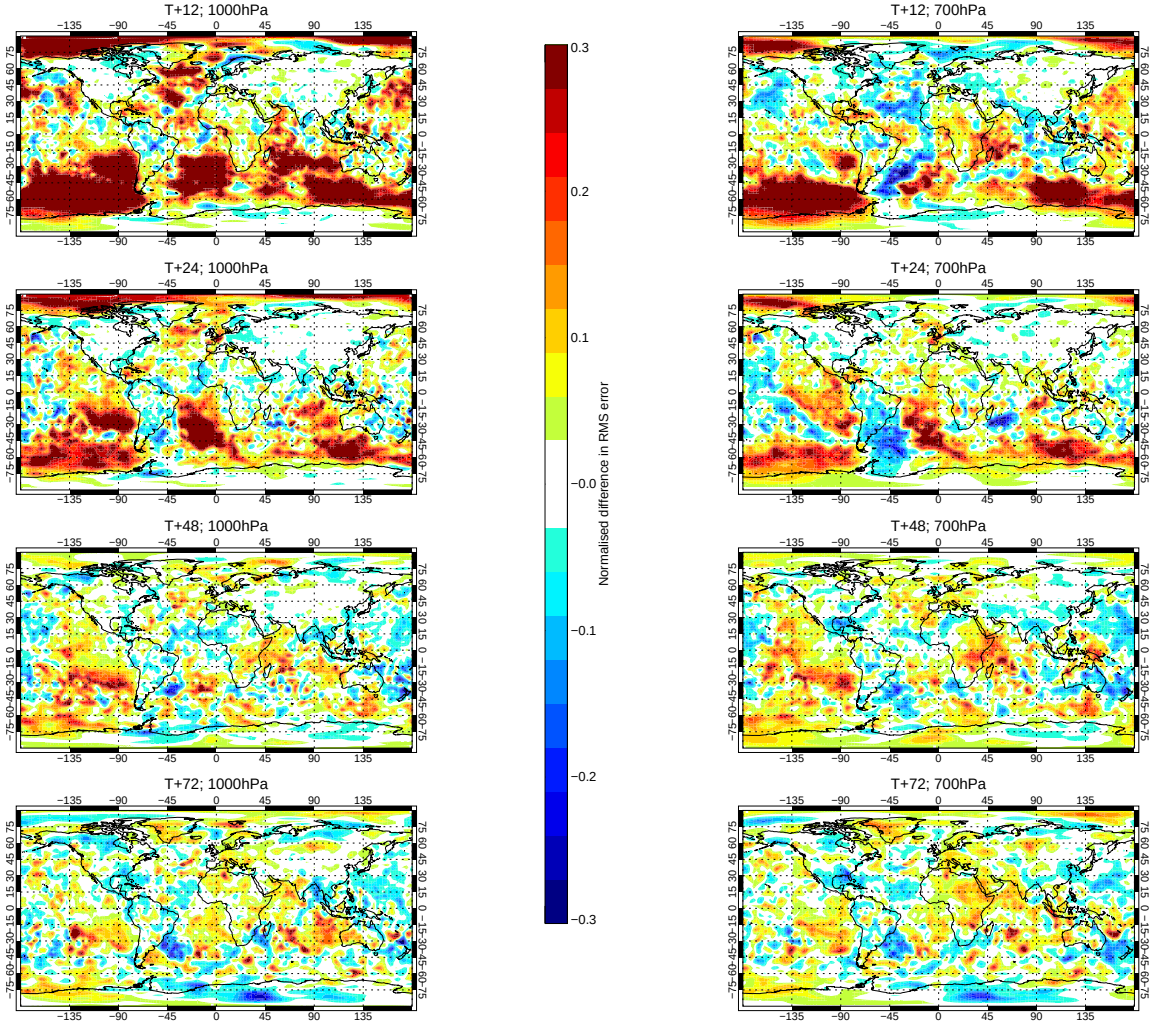


Figure 42: Normalized RMS forecast error difference between Baseline-2 and Control-2 experiments for 1000 hPa geopotential height (left) and 700hPa geopotential height (right). Positive values indicate positive impact of the assimilated buoy data. Panels show forecast range of 12, 24, 48 and 72 hours (from top to bottom). Forecast verification is against operational analyses; verification period is 7/12/2008-24/01/2009.

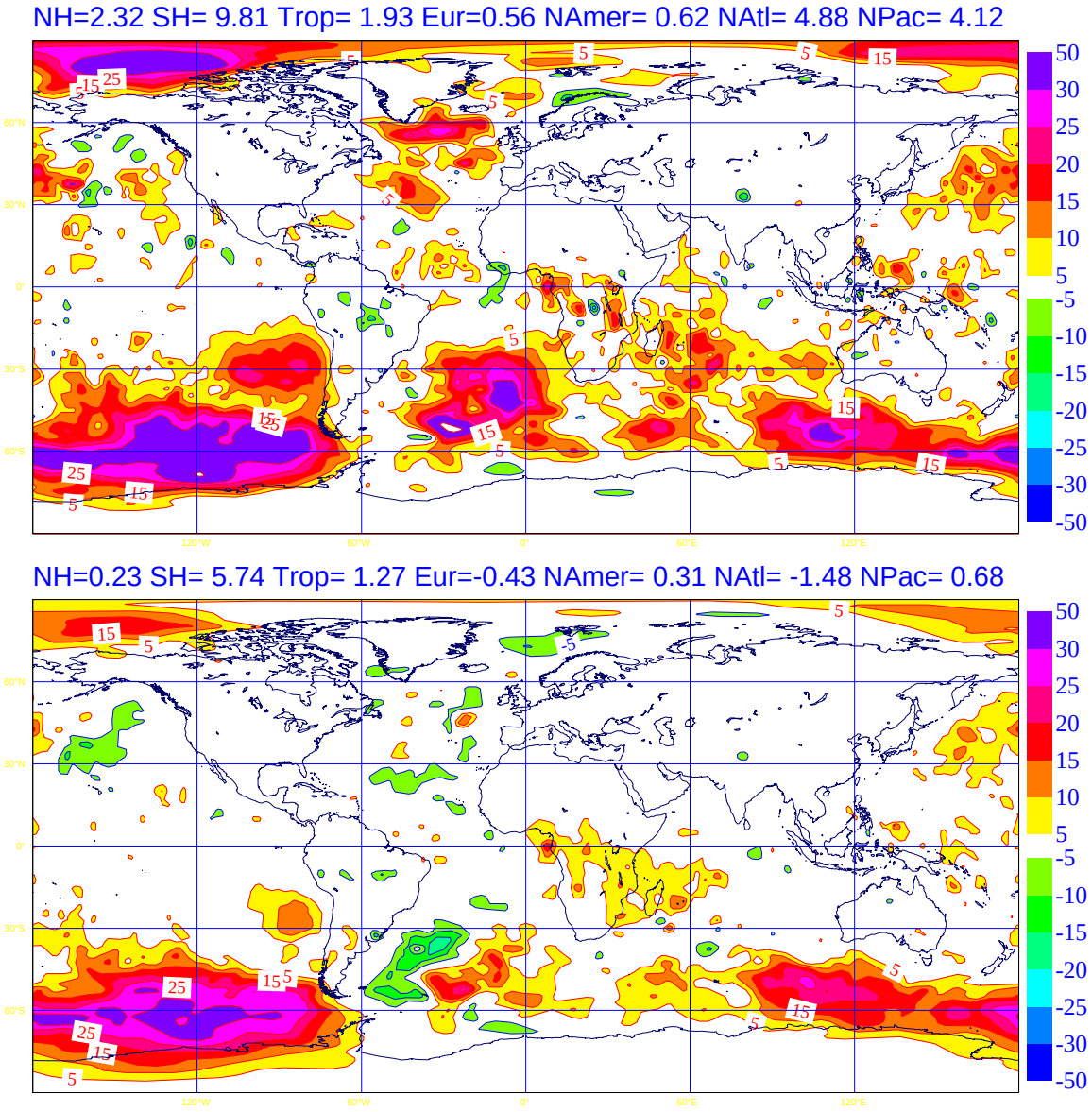


Figure 43: RMS forecast error difference between Baseline-2 and Control-2 experiments for 1000 hPa geopotential height (top) and 700hPa geopotential height (bottom). The header of the maps shows the mean RMS error reduction over 7 different regions of the globe in $m**2/sec**2$. Positive values indicate positive impact of the assimilated buoy data. Maps show forecast range of 12h. Forecast verification is against operational analyses; verification period is 7/12/2008-24/01/2009.

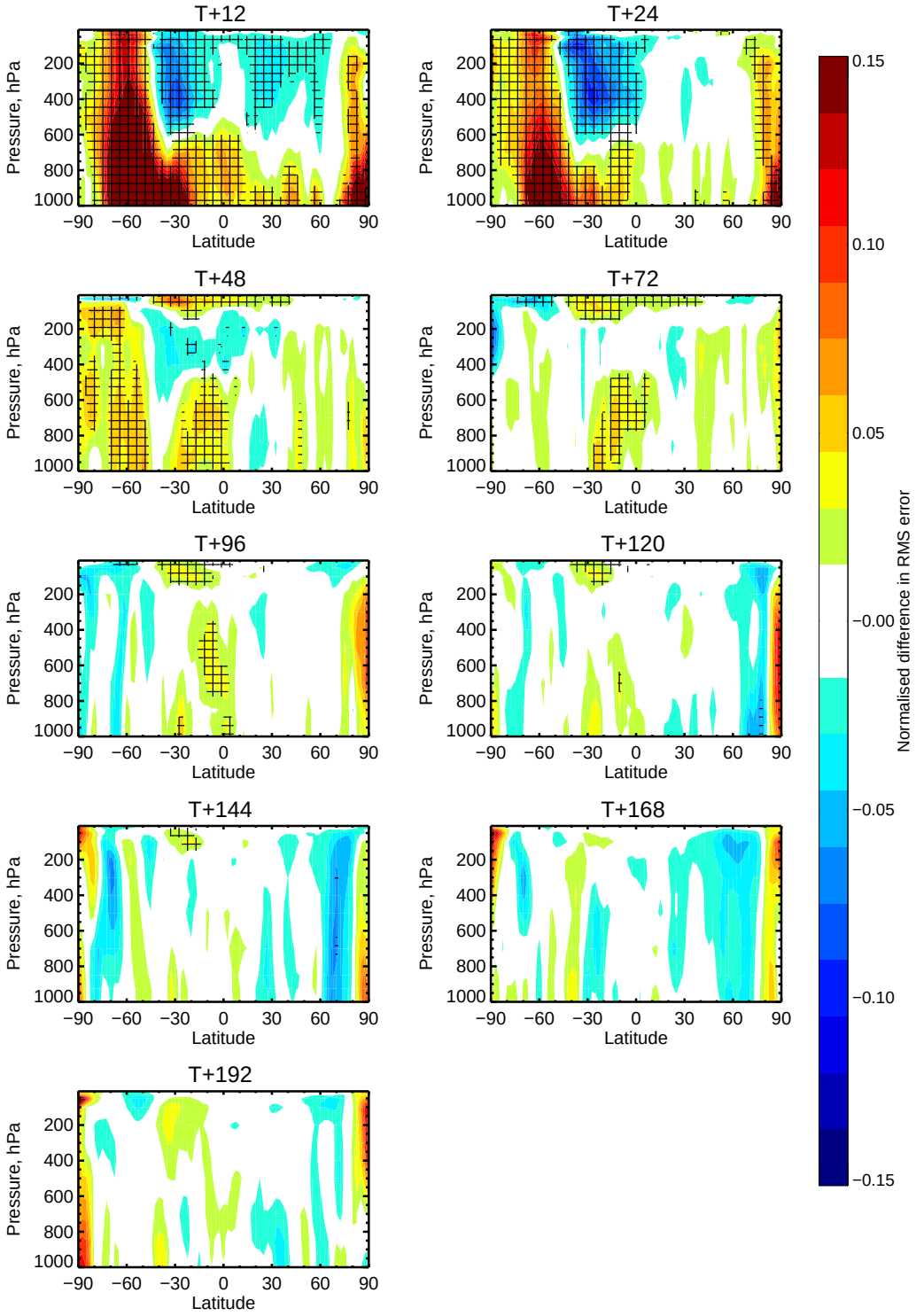


Figure 44: Zonal cross-sections of normalized RMS forecast error difference between Baseline-2 and Control-2 experiments for geopotential height. Positive values indicate a positive impact of the assimilated buoy data. Panels show forecast range of 12, 24, 48, 72, 96, 120, 144, 168 and 192 hours. Crosses indicate where scores are statistically significant to the 95% level. Forecast verification is against operational analyses; verification period is 7/12/2008-24/01/2009.

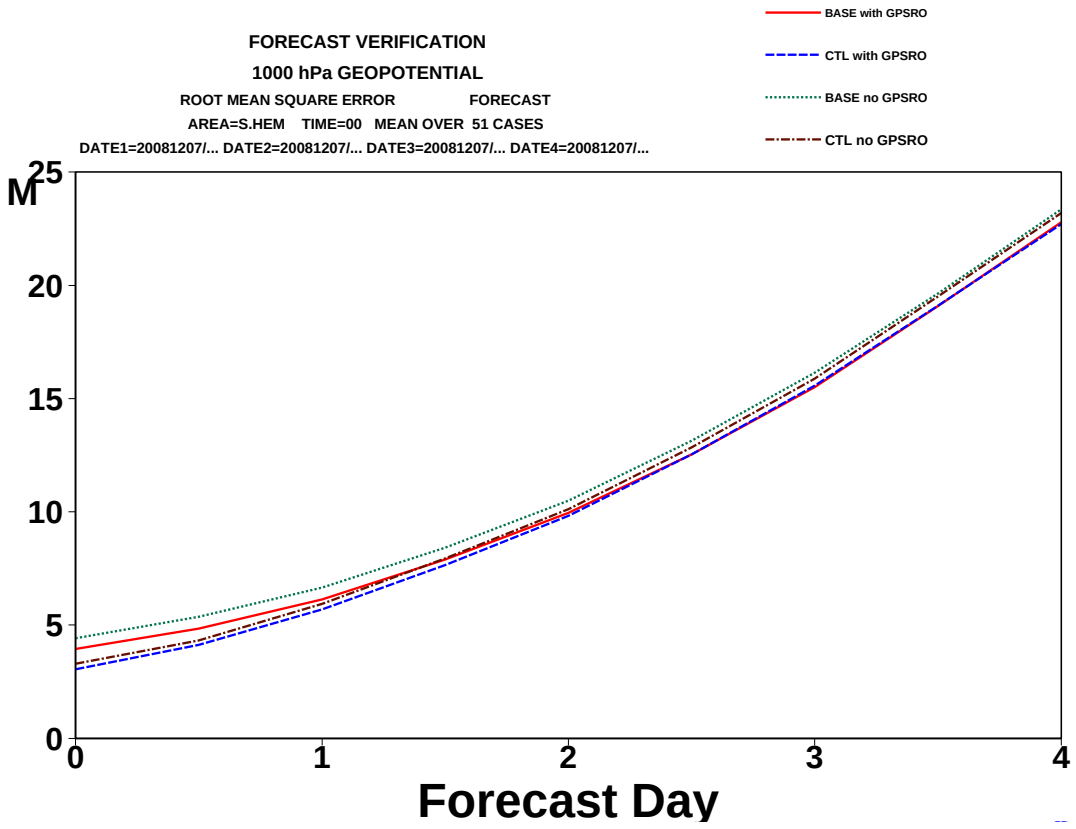


Figure 45: RMS error curves of 1000hPa geopotential height for the two baseline and two control experiments for the Southern hemisphere. Verification period is 7/12/2008-26/01/2009.

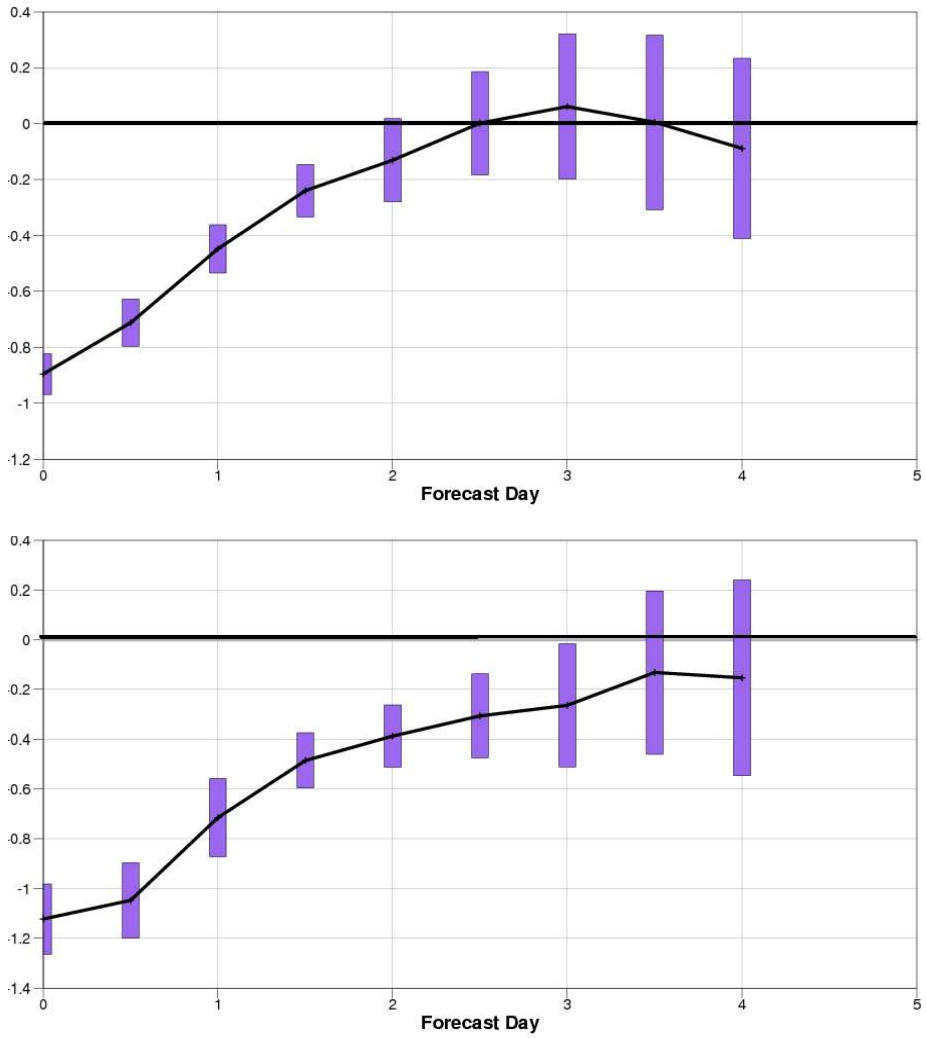


Figure 46: RMS error difference curves between Control and Baseline with (top) and without (bottom) GPSRO data assimilated for 1000hPa geopotential height for the Southern hemisphere. Verification period is 7/12/2008-26/01/2009. Units are in m. Vertical bars represent significance of the results on 90% confidence level.

5.3.2 Experiments with thinned buoy data

Two experiments (one with and one without GPSRO assimilation) have been performed for the same period with thinned buoy data and reduced temporal frequency of VOS reports. These experiments simulate the in situ measurement coverage that had been available before the North-Atlantic buoy network was updated within the framework of the E-SURFMAR program. The purpose of these experiments is to assess the impact of the enhanced in situ North Atlantic observation network on the performance of the IFS model. The RMS forecast error difference maps over the North Atlantic region between the experiments with thinned North-Atlantic in situ measurements and the control in the case when GPSRO data are not assimilated can be seen on Figure 47 for 1000hPa geopotential height for forecast time ranges of 12h (top), 24h (middle) and 48h (bottom). Again, the header of the maps shows the mean RMS error reduction over 7 different regions of the globe in $m**2/sec**2$. It can be seen that over the North-Atlantic region the additional data have a moderate positive impact on the RMS errors. This positive impact is slightly amplifying with forecast range up to 24-48h and some of the impact is propagating towards Europe. It is surprising however, that some negative impact can be detected at the Scandinavian coasts, one of the areas where the thinning completely removed the buoys. It has to be noted that the same impact can be observed in the baseline scenario on Figure 43. The same RMS difference map with GPSRO data assimilated can be seen on Figure 48, only for the 12h forecast range. It is very similar to what has been seen without GPSRO data, the same dipole structure can be recognized at the Scandinavian coasts and the same strong positive impact of the extra data South of Greenland. Figure 49 shows for the North-Atlantic area the Control versus Thinned RMS error difference curves with (top) and without (bottom) GPSRO data assimilated. The vertical bars represent the significance interval of the results on 90% confidence level. As it can be seen, the positive impact of the extra buoy data is significant up to 24 hours when GPSRO data are present and up to 48 hours without GPSRO data. The small negative impact on other time ranges is statistically insignificant. The impact of the additional buoy data on precipitation forecast as compared against Western Europe SYNOP measurements has been examined. The results showed (not shown here) that in terms of precipitation the differences are statistically insignificant and they depend on the precipitation threshold values used in computing the statistics.

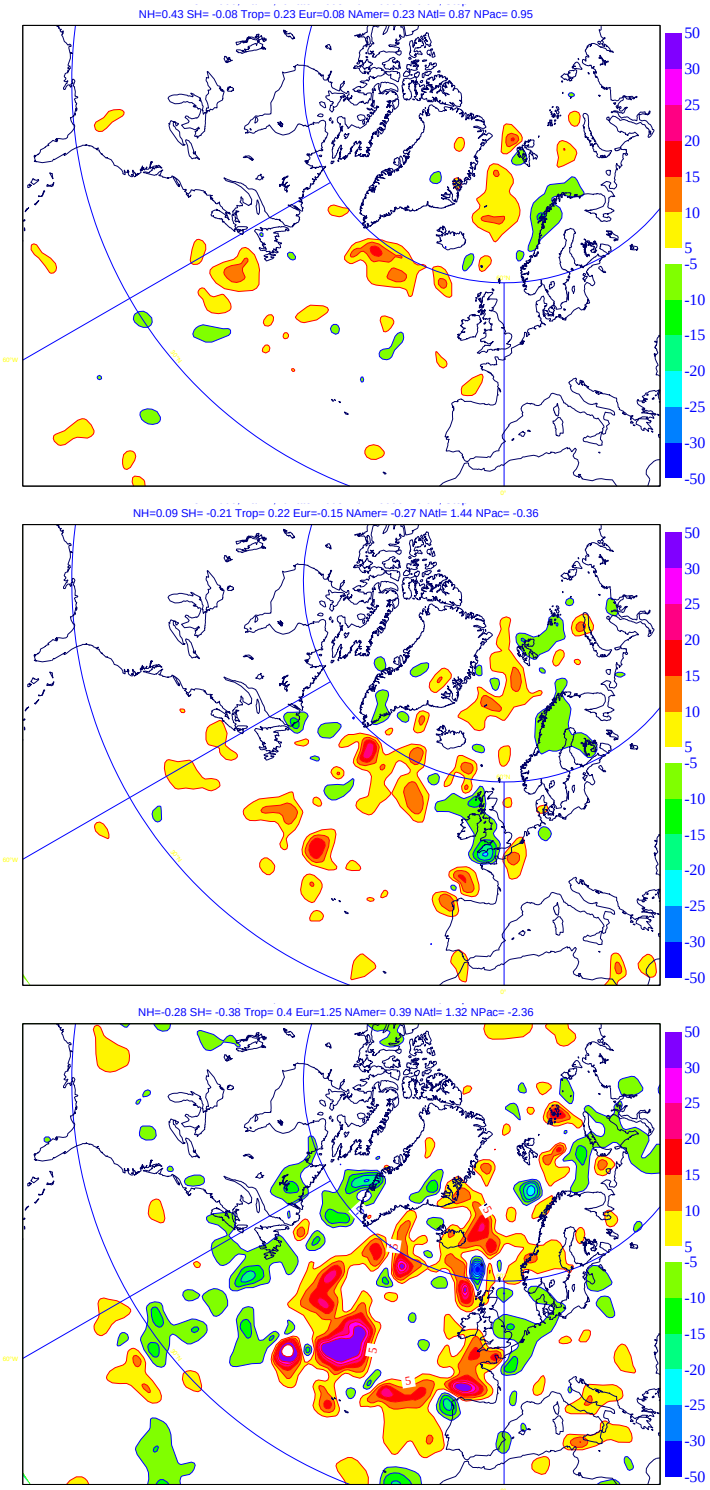


Figure 47: RMS forecast error difference maps over the North Atlantic region between Thinned-2 and Control-2 experiments for 1000 hPa geopotential height for time ranges of 12 (top), 24h (middle) and 48h (bottom). The header of the maps shows the mean RMS error reduction over 7 different regions of the globe in $m**2/sec**2$. Positive values indicate positive impact of the assimilated additional buoy data. Forecast verification is against operational analyses; verification period is 7/12/2008-26/01/2009.

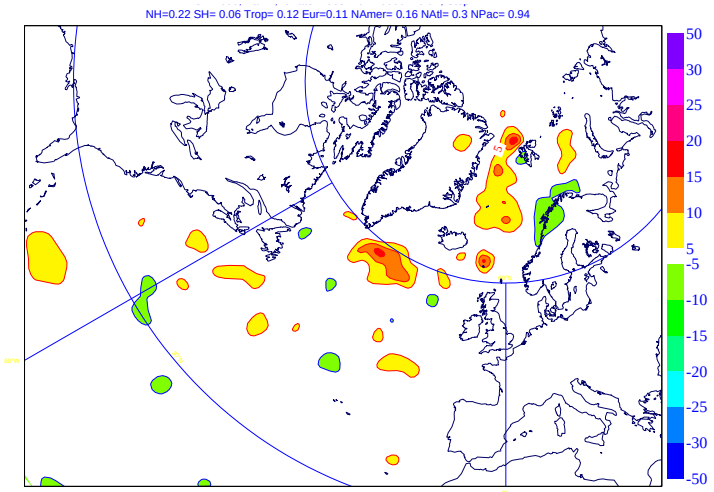


Figure 48: Same as Figure 47, but with GPSRO assimilated and only 12h forecast range.

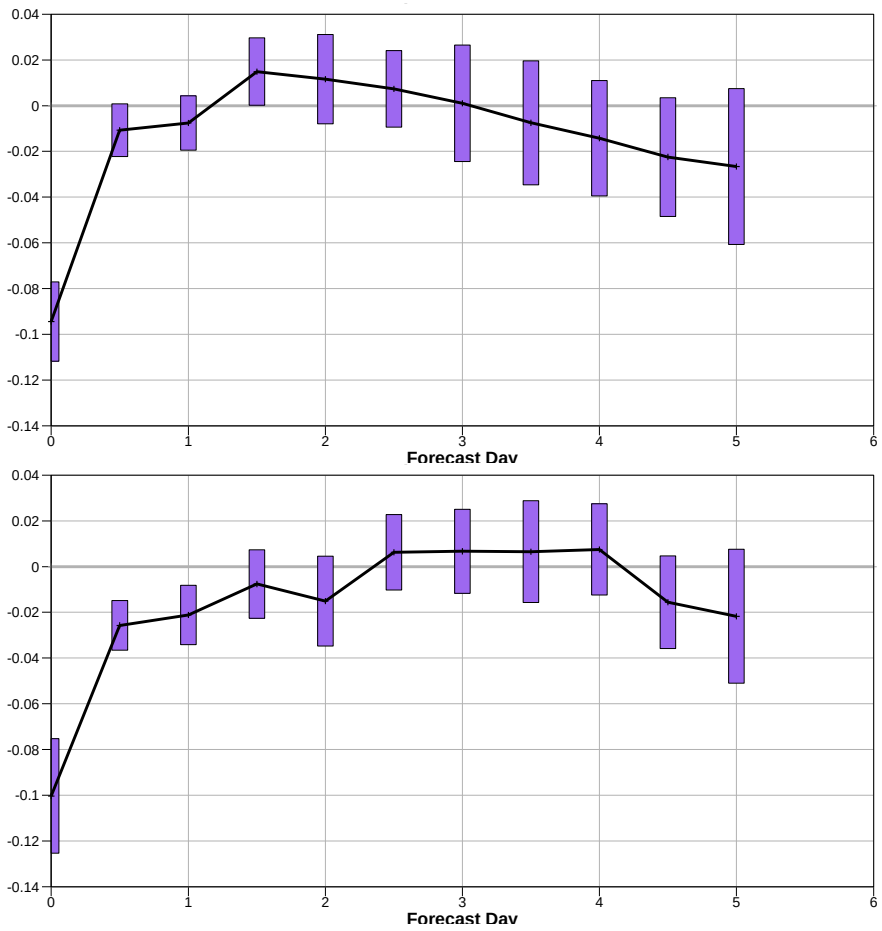


Figure 49: RMS error difference curves between Control and Thinned with (top) and without (bottom) GPSRO data assimilated for 1000hPa geopotential height for the North-Atlantic area. Verification period is 7/12/2008-26/01/2009. Units are in m. Vertical bars represent significance of the results on 90% confidence level.

It has been seen that the impact of the data thinning on forecast performance is much smaller than the impact of full denial of buoy data, but still significant, at least in the short range (up to 1-2 days). This impact can be larger, however, in cases of severe weather events. This question has been examined for two North-Atlantic intensive winter storms Klaus (January, 2009) and Xynthia (February, 2010). The former storm happened within the examined period while for Xynthia the above experiments have been rerun with a 5 day warm-up period including the cyclogenesis. Figure 50 (left and middle) shows the 48h mean sea level pressure forecast for storm Klaus with the thinned and full buoy network, respectively, both without GPSRO data assimilated. The right panel shows the verifying control analysis. It can be seen that both forecasts underestimate the depth of the low, but the one using more buoy data is more accurate both in terms of position and depth of the low. For these 48h forecasts, this impact is somewhat larger than in the case when both experiments run with assimilating GPSRO data (not shown here). Figure 51 shows at the same verification time the 96h forecasts in the six different scenarios (control, thinned and baseline without GPSRO on the top panels and the same with GPSRO data on the bottom panels). In this case the differences of the different scenarios are larger and it is seen that phase error of the predicted low is getting smaller when more buoy data are used. At the same time with more buoy data the low is slightly deeper, which is also improving the forecast.

A similar comparison has been performed for the storm Xynthia which hit Spain and Southern France one year later in 2010. In this case the model has predicted the storm track more accurately than in the case of Klaus and the differences between the performance of the different data usage scenarios are smaller than in the case of Klaus. It has to be noted that Klaus and Xynthia have quite different storm tracks and the expected impact of the additional buoy observations naturally strongly depends on the storm tracks. Xynthia moved from the sub-tropical North-Atlantic to North-East direction and both hit the continent near the Bay of Biscay. However, if one compares 96h forecasts of Xynthia starting from the analyses prior to the time of cyclogenesis, some differences between the different scenarios can be identified. Figure 52 shows the 96h mean sea level pressure forecasts valid at 27, February, 2010, 00UTC of the four different scenarios (top two panels without GPS data, middle two panels with GPS data) and the verifying analysis (bottom figure). It can be seen that the assimilation of the additional buoy observations results in deepening the low (left panels show deeper lows than right) and it makes the forecast more accurate as compared with the verifying analysis. On the other hand it is interesting to see that the assimilation of GPSRO data introduces a phase error (top figures compared to middle and bottom ones), i.e. in this case the model propagates the storm slightly faster to the North-East than in reality or in the experiments without GPSRO data.

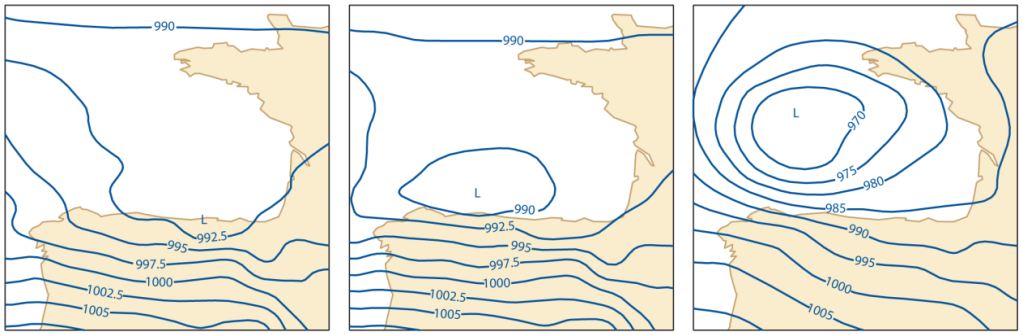


Figure 50: Mean sea level pressure 48h forecast field valid at 00UTC 24 January, 2009 (storm Klaus) from analyses without GPSRO data and with thinned North-Atlantic buoy network (left), with full North-Atlantic buoy network (middle) and the corresponding verifying control analysis (right).

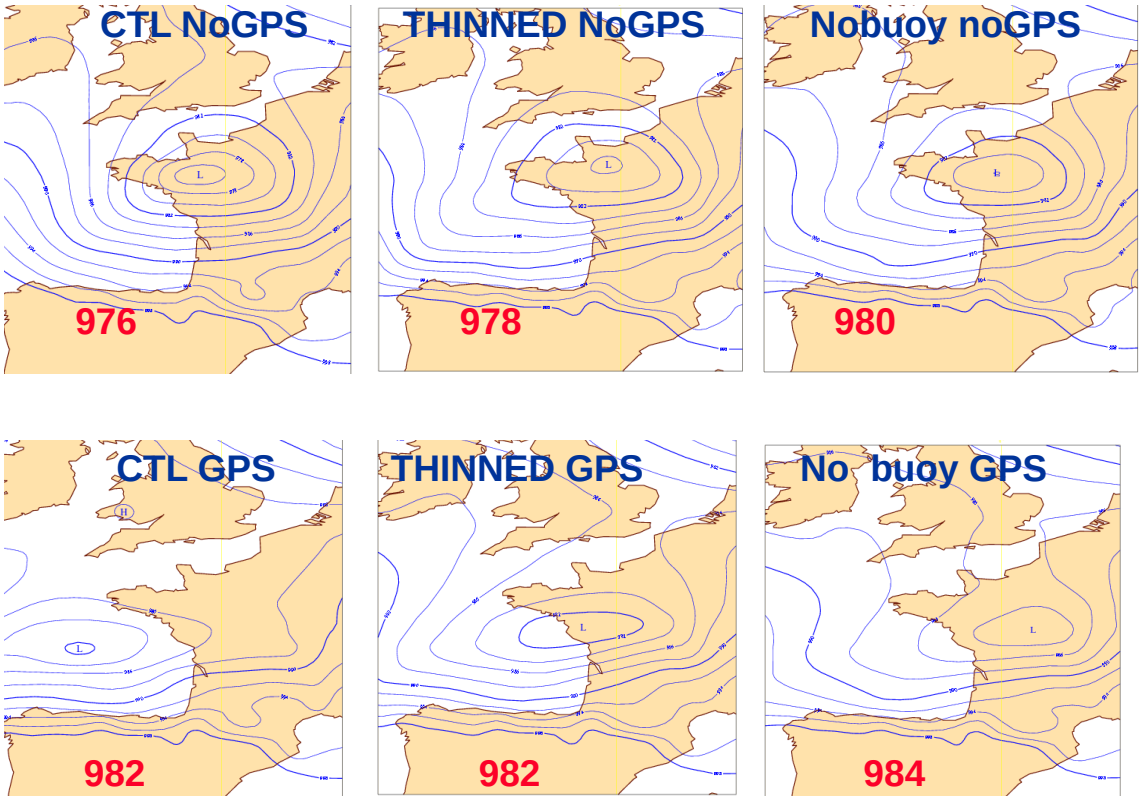


Figure 51: Mean sea level pressure 96h forecast field valid at 00UTC 24 January, 2009 (storm Klaus) from analyses without (top) and with (bottom) GPSRO data, with assimilating the full buoy network (left), reduced buoy network (middle) and no buoy data at all (right). The red numbers on each panel denote the forecasted depth of the low in hPa. The corresponding verifying analysis in the right panel of Figure 50

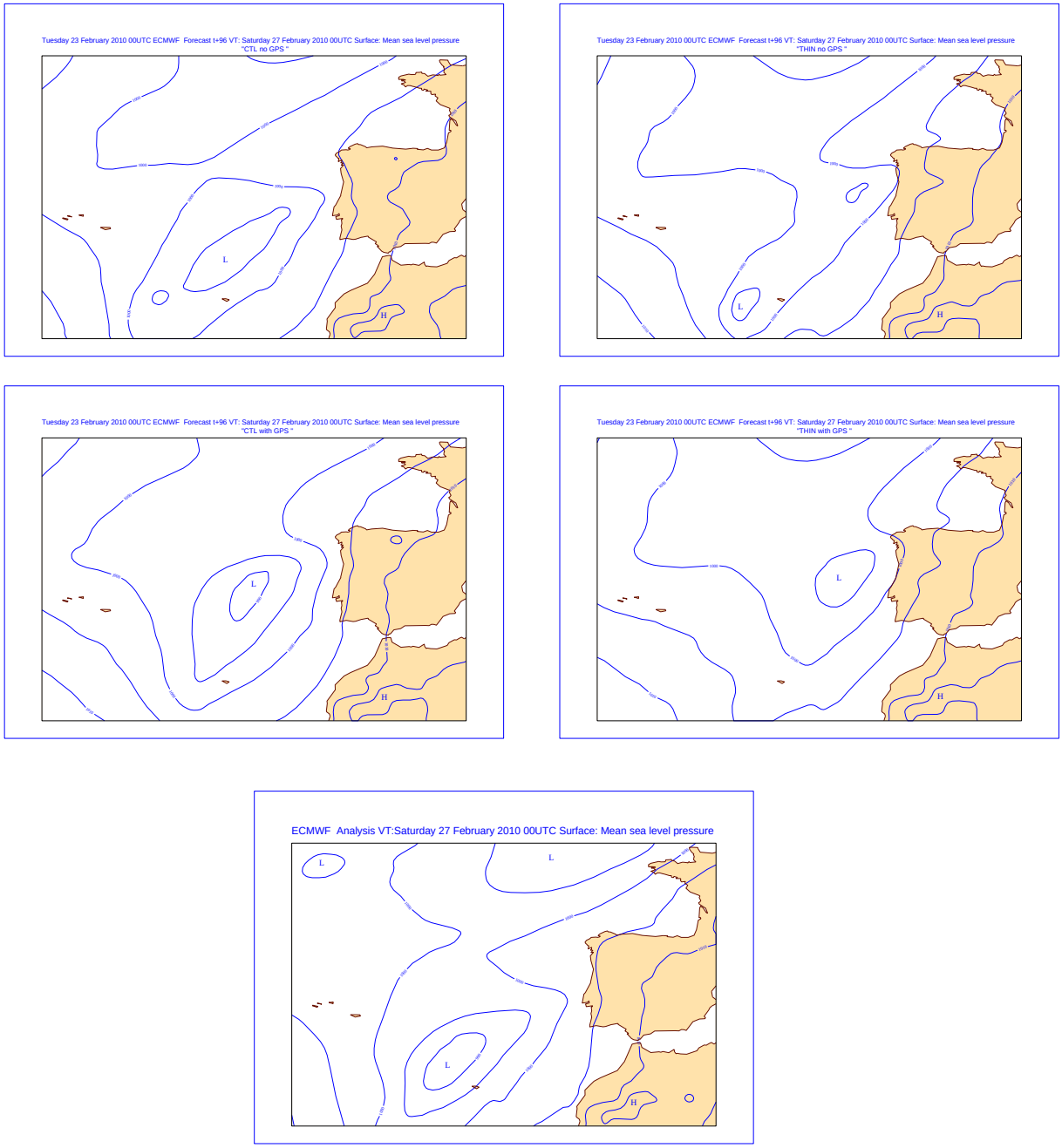


Figure 52: Mean sea level pressure 96h forecast field valid at 00UTC 27 Februaryy, 2010 (storm Xynthia) from analyses without (top) and with (middle) GPSRO data, with assimilating the full buoy network (left) and reduced buoy network (right). The bottom panel shows the verifying analysis.

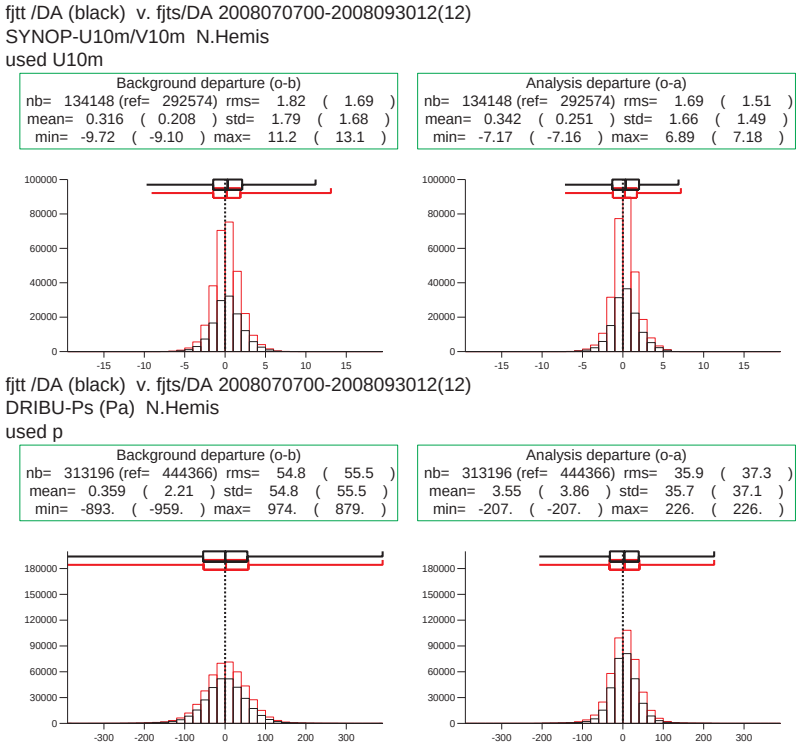


Figure 53: Fit between model first-guess (left) and analysis (right) and observations in Northern hemisphere SYNOP 10m wind data (top) and buoy surface pressure observations (bottom) (red curves control experiment, black curves denial experiment). Statistics were generated from period 7/7/2008-30/09/2008.

5.3.3 Follow-on investigation of the buoy/ship impact study

It has been clearly shown in the 2 month long impact study that the total denial of buoy and non-synoptic time VOS measurements has a large detrimental impact on the forecast performance. It has also been shown that in some extreme weather events partial removal of these measurements, i.e. data thinning to a reduced network representing the data resolution before the "E-SURFMAR era", can have a remarkable impact on the individual forecasts. However the length of experiment was not sufficient to draw more general conclusions from the partial data thinning experiments. Therefore the study has been extended to a 6 month comparison of the performance with the thinned and full data coverage. The extended experiments were performed without GPSRO assimilation because of some considerations described in the previous sections and the period covered the second half 2008. The difference between the two experiments (control and denial experiments) is again only in terms of data usage (buoy and VOS data). Figure 53 illustrates the data usage difference for these data types over the first 3 months of the experiments. The most visible differences are seen for 10m wind and surface pressure assimilation. 10m winds are used in our model only over the sea, therefore the removal of non-synoptic time VOS data results in a more than 50% loss of data over the Northern Hemisphere (top panel). The Atlantic thinning of the buoy network removes one quarter of the asimilated buoy surface pressure data over the Northern Hemisphere (bottom panel). In the earlier, two month experiment it has been shown that some positive impact of the additional data can be identified on 24-48h geopotential forecast. Figure 54 shows the same forecast error reduction now only for 24 and 48h time ranges derived from the 6 month experiment pair. Note that the contour scaling is slightly different from Figure 47 because in the longer experiment the impacts became smaller. The main signal is that a moderate positive impact of the additional data can be ob-

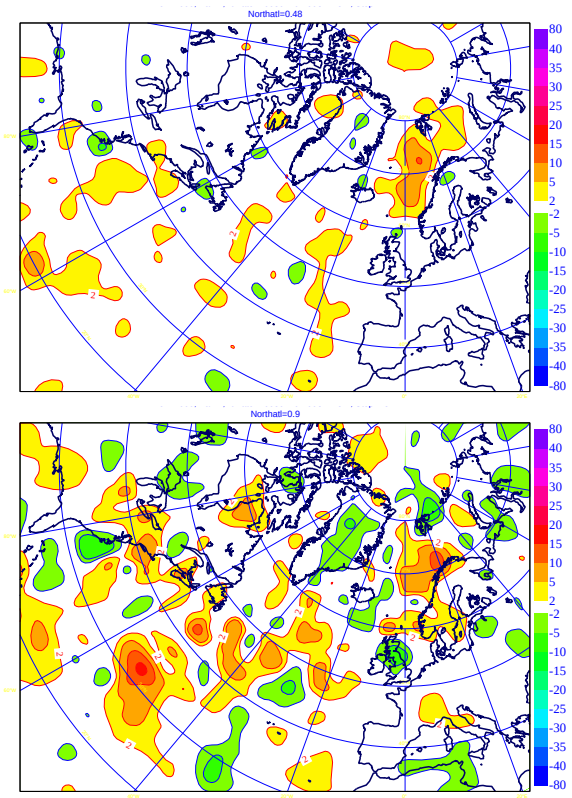


Figure 54: RMS forecast error difference maps over the North Atlantic region between Thinned and Control experiments for 1000 hPa geopotential height for time ranges of 24 (top) and 48h (bottom). The header of the maps shows the mean RMS error reduction over the North-Atlantic region in $m**2/sec**2$. Positive values indicate positive impact of the assimilated additional buoy data. Forecast verification is against operational analyses; verification period is 10/07/2008-25/12/2008.

served, especially at the 48h forecast range along a rather well defined North-East to South-West bend in the North Atlantic region spreading from the coastal region of Scandinavia down to the Tropical Atlantic Ocean. The positive impact at the Scandinavian coasts is the largest at the 24h range and it coincides with additional buoys introduced by E-SURFMAR for this area. It has to be mentioned that the earlier seen negative impact over Scandinavia is not present in the 6 month experiment. Figure 55 shows the RMS error difference between control and denial (negative values mean positive impact of the additional data) for 1000hPa (top) and 850hPa (bottom) geopotential (left) and temperature (right) over the North-Atlantic domain as verified against operational analyses. It is clearly seen that up to 48-72 hours the impact of extra buoy/VOS data is significantly positive on 95% confidence level. The overall impact of extra buoy/VOS data can be summarized on the scorecards presented in Figure 56. The scorecards show the RMS error differences between control and buoy/VOS denial experiment for different variables, areas, vertical levels and forecast ranges. Green color represents positive forecast impact of the additional data, triangles represent significance of the impact, the bigger the triangle, the stronger the significance. The figure spans the 12h-144h forecast range and statistics were dervied from 00 and 12UTC forecast from the period 07/07/2008-31/12/2008. It is very clear that the impact on every variable and level is positive or neutral. Interestingly, over the whole Northern Hemisphere the impact is stronger than just over the Northern Atlantic area. It is also surprising that the impact spans the whole troposphere and it goes well to the medium-range (up to 7 days).

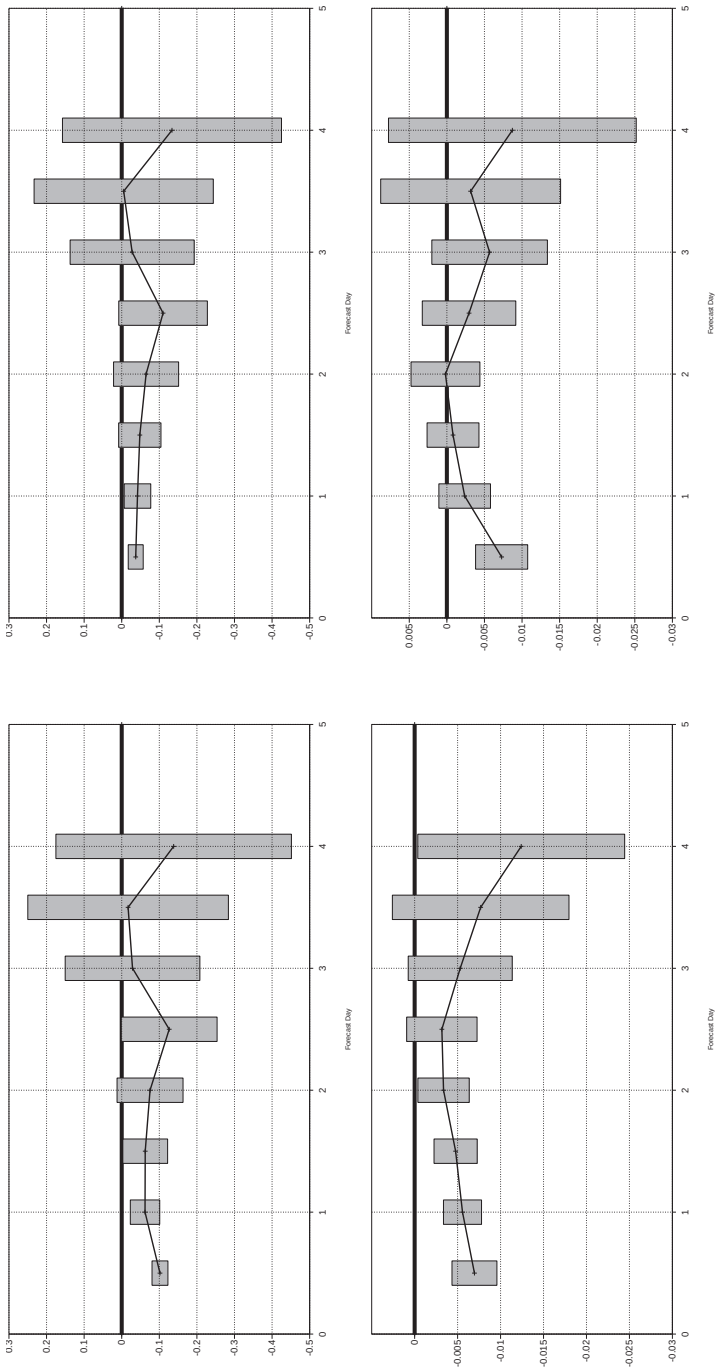


*Figure 55: RMS error differences between control and buoy/VOS denial experiment for geopotential height (left in m**2/sec**2) and temperature (right in K), for 1000hPa (top) and for 850hPa (bottom). Vertical bars represent significance interval on 95 percent confidence level. Verification is derived from 00 and 12UTC forecasts and it is performed against oparational verifying analysis. Verification domain is North-Atlantic region. Verification period is 07/07/2008-31/12/2008.*

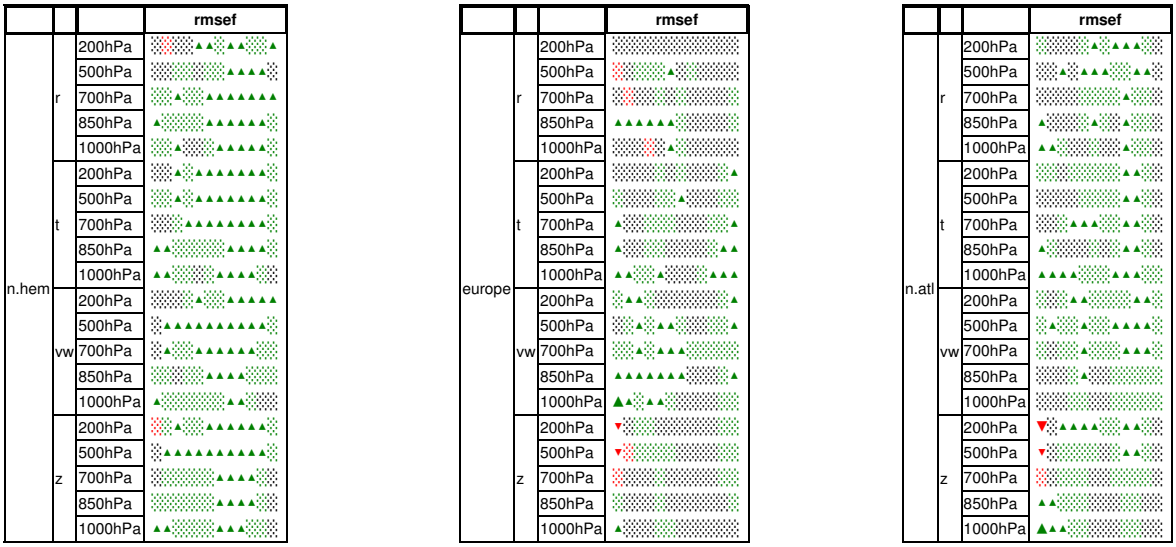


Figure 56: Scorecard of RMS error differences between control and buoy/VOS denial experiment for different variables, areas, vertical levels and forecast ranges. Green color represents positive forecast impact of buoy/VOS data. Triangles represent significance of the impact, the bigger the triangle, the stronger the significance. Figure spans the 12h-144h forecast range. Statistics are dervied from 00 and 12UTC forecast from the period 07/07/2008-31/12/2008. Domains, levels and variables are indicated on the figure.

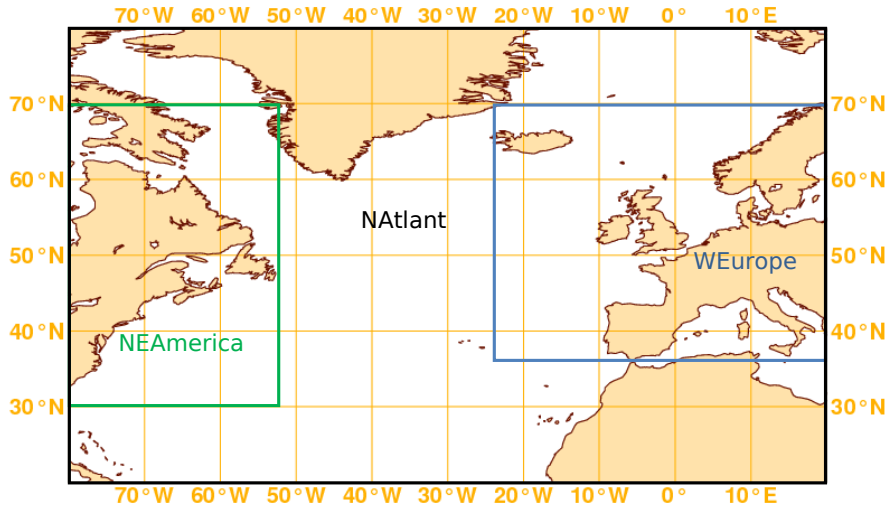


Figure 57: Location of the 3 domains used for the precipitation verification in the buoy/VOS denial experiments. The domain 'NAtlant' covers the whole area.

The impact of the data denial on precipitation forecast performance has also been analyzed following the same methodology as in subsection 5.2. The verification period for the comparison is 1 July - 31 Dec 2008, the first 5 days have been left out of the verification to exclude spin-up effects. The definition of domains is shown in Figure 57 , the location of available surface stations during the verification period is shown in Figure 58 . The NE-America domain contains 180 stations, the W-Europe domain 1000 stations, and the full domain 1300 stations. Witholding the total buoy/VOS network has little effect on the SEEPS score for the 12UTC run (Figure 59, lower panels). Differences between experiments are non-systematic, and within the sampling uncertainty. A small systematic effect is seen for the 00Z run (Figure 59, upper panels), where the denial experiment scores slightly worse than the control both in Western Europe and Northeast America (only Western Europw shown here). This is most pronounced on forecast days 3-5. Verification has also been performed using the Equitable Threat Score (ETS) and the Peirce Skill Score (PSS). At a threshold of 1 *mm*, the denial experiment scores slightly worse than the control for the 00UTC run, but slightly better than the control for the 12UTC run. At a threshold of 5 *mm* the differences between denial and control are even smaller than for 1 *mm*, and not significant. All these impacts are summarized on the verification scorecard on Figure 60. On the verification scorecard red colours indicate the positive impact of the extra observations. For each domain and run, SEEPS, ETS, and PSS scores are presented and the difference between experiment and control is indicated by color codes. For the higher thresholds of 5 and 10 mm only the ETS is shown, because the PSS asymptotically reduces to the POD as events become less frequent. It can be seen that overall there is no clear, systematic signal, but there is a tendency for the 00UTC denial run to deteriorate from day 3 onwards compared to the control. For 12UTC, days 3-5 are, if anything, slightly improved in the denial, but for higher thresholds the deterioration dominates here as well.

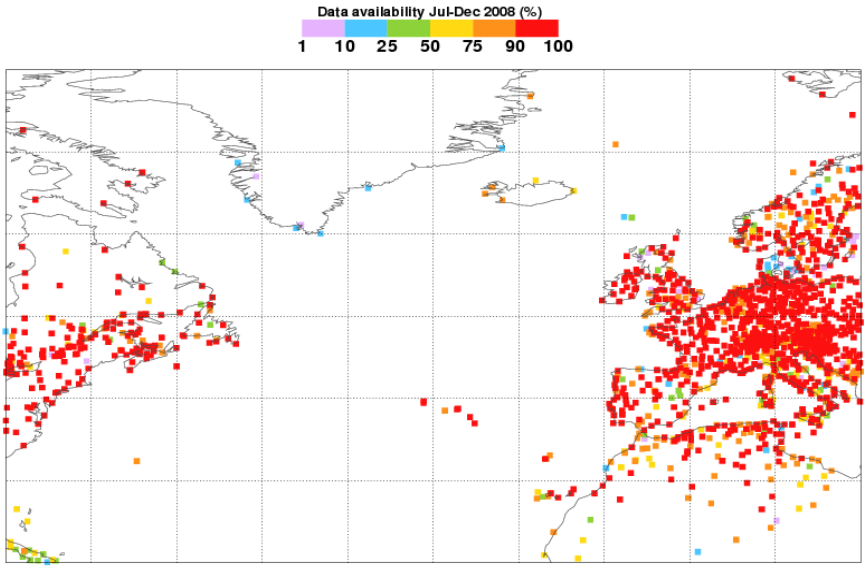


Figure 58: Geographical distribution of surface stations used for precipitation verification in the buoy/VOS denial experiments.

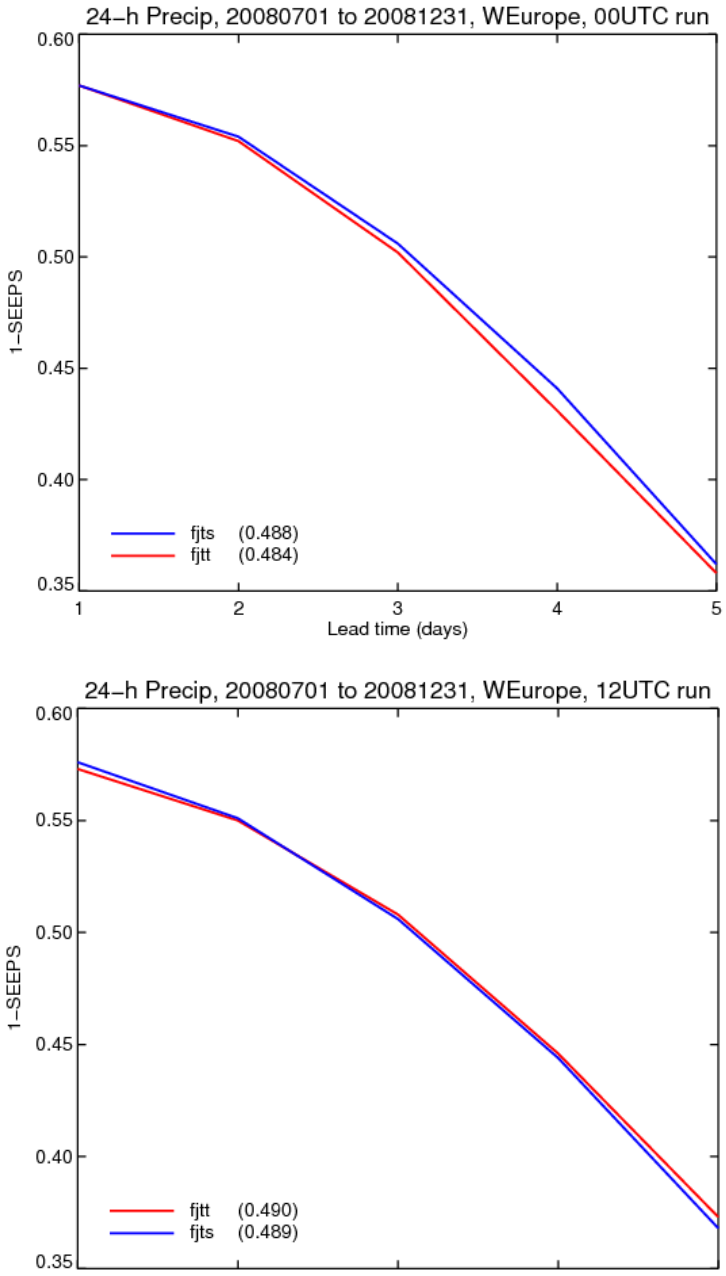


Figure 59: $1 - SEEPS$ *for the 00 UTC (top) and the 12 UTC run (bottom) for Western Europe as a function of lead time for the control (blue) and the buoy/VOS denial experiment (red).*

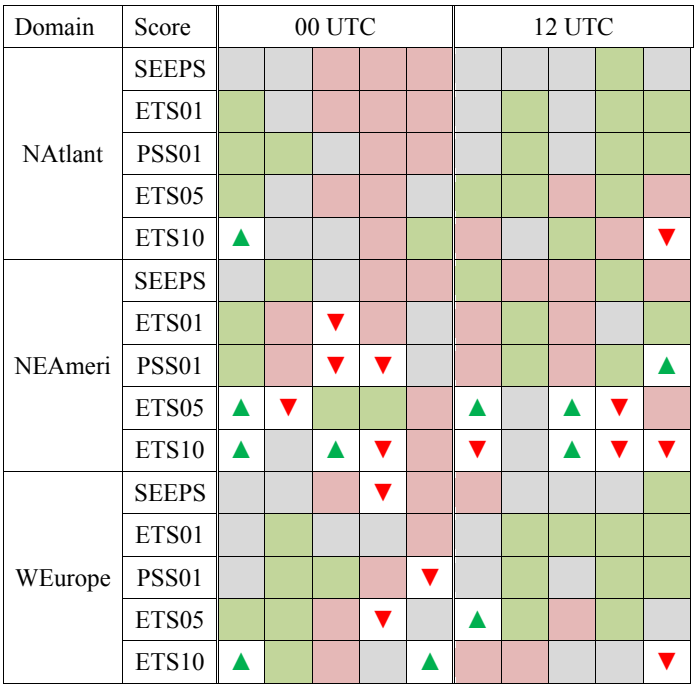


Figure 60: Summary of verification results for buoy/VOS denial experiment vs control. Columns show individual differences for forecast days 1-5 for each model run. Red (green) triangles indicate a positive (negative) score impact of the additional data, difference of 0.01 or more, red (green) shading indicates a positive (negative) score difference of 0.003-0.009, grey shading indicates a score difference of 0.002 or less. Numbers next to ETS and PSS denote thresholds in mm.

5.3.4 Summary

In this work package the impact of marine buoy surface observations and VOS observations has been examined by running observation system experiments (OSEs). A special emphasis has been put on the impact of North-Atlantic marine buoys. A two-month OSE from December 2008 to January 2009 has been performed to assess the impact of the extra surface pressure measurements that were introduced during the last decade in the framework of the EUCOS E-Surfmar program. The control experiment used the operational observing system. The OSE has been run with the standard 4D-Var system using T511 horizontal resolution and 91 vertical levels. The impact of the global denial of buoy data has been also investigated. Both the control and the denial experiments have been repeated without GPSRO radio occultation data, since GPSRO assimilation has been recently proven to constrain the surface pressure analysis. Later the experiment pair of partial denial without GPSRO data has been repeated for a 6 month period to see the significance of the relatively small impact of partial data denial. The forecast impact of buoy and GPSRO data has been examined in two severe winter storm cases (storms Klaus and Xynthia), that recently affected Europe. The following conclusions can be drawn form the performed experiments:

- The impact of buoy surface pressure observations on the forecast performance is large, especially in the lower atmosphere and it is more expressed over the Southern Hemisphere.
- This impact is more evident when GPSRO data are not assimilated: significant impact lasts 2-3 days with and 4 days or more without assimilation of GPSRO data.
- The comparison of the surface pressure forecast score impact of buoy and GPSRO data shows that buoy data are more important until 24h while from 48h onwards GPSRO data clearly contribute more to the forecast quality. When going higher in the atmosphere GPSRO data become more and more dominating over buoy data.
- Additional buoy data introduced at the North-Atlantic area within the E-SURFMAR program prove to locally improve surface pressure forecast scores but this impact is moderate and it lasts up to 24-72h.
- The impact of the additional buoy data can be clearly seen in extreme weather events.
- To detect impact of the additional data on precipation scores is diffiult, however some moderate positive impact can be identified.

5.4 ASAP impact study

An experimentation was performed in order to study the impact of the E-ASAP shipborne radiosonde network in the forecasts of the IFS model. The list of EUCOS/ASAP maintained radiosondes on ships was considered from the EUCOS webpage and 18 of the 19 ASAP ships were blacklisted in the experiment (the last one, which is ASIS01, a temporary land station operated by Vedurstofa Islands was not excluded). The impact study was carried out for a 3-month period during July-September, 2011. The applied IFS model cycle was cy37r3, the horizontal spectral resolution was T511.

Figure 61 shows the distribution of the ASAP soundings over the Northern Atlantic. Note that the figure indicates the one-month accumulation of data. Figure 62 gives a hint about the relative quantity of the excluded data with respect to all the available sonde measurements over the Northern Hemisphere (for temperature for the entire 3 months period). It can be seen that the data denial results in a decrease of 1-1.5 % of data (the larger values can be found near to the surface) and moreover the denied data don't

have an influence on the O-G (observation minus guess) and O-A (observation minus analysis) statistics. Having a look on individual dates, it can be seen that on average there are 5-6 extra radiosondes available at every 12 hours for inclusion into the data assimilation process.

The results of the control and denial experiments were verified against the IFS operational analysis, against their own analysis and against observations, respectively. Mostly, a North-Atlantic domain was considered, but also various verification domains over Europe were used.

The best overall judgement of the results can be gained if the so called scorecards are visualised. The scorecards summarise the verication differences (indicating its sign and significance) between the two experiments in terms of selected domain (North Atlantic, Europe for instance), altitude (850 hPa, 700 hPa, 500 hPa, 200 hPa and 100 hPa), verification measure (RMSE or anomaly correlation), prognostic variables (wind, humidity, temperature, geopotential). Similar scorecards were produced for precipitation scores as well. These diagnostics make a separation between the 00 UTC and 12 UTC integrations and use several precipitation scores (SEEPS, ETS01, PSS01, ETS05, ETS10) for different domains (North Atlantic, NE America and West-Europe). See more details at Figure 57 and Figure 58 and at the corresponding text of subsection 5.3.3.

Hereafter while analysing the results based on the scorecards the changes will be considered as positive (negative), when the control is providing better (worse) forecasts than the denial experiment, i.e. the additional data available at the control makes the forecast better.

The most visible positive differences are for the North-Atlantic domain (Figure 63) at the beginning of the model integration. It is mostly apparent for 850 hPa temperature (for verification against own analysis) and wind field at 850, 700 and 100 hPa (against operational analysis). However, the confidence on these positive results is undermined by the discrepancy between the scores using different analysis (operational or own) as the truth and also by the generally small significance of the differences. For a European domain (Figure 64) the differences between the two experiments are further decreased and even some significantly negative differences appear for the scores against own analysis for the 4-5 days forecast ranges. While considering European and Northern African observations as verification reference (Figure 65) negative scores are visible for geopotential below 500 hPa (mostly at 4-5 days, but also at the beginning of integration in terms of RMSE).

As far as the precipitation scorecard (Figure 66) is concerned the results are rather negative. Particularly the 12 UTC runs for higher precipitation thresholds after three-days of forecast. Degradation is visible for the NE American domain as well, while for Western Europe the scores are rather insignificantly negative.

As a summary it can be said that there are some slightly positive impacts (the additional ASAP soundings improve the forecasts) found at the first two forecasting days and slightly negative afterwards. The precipitation results show small, but clear negative signal, i.e. the denial experiment is performing better than the control (especially for the 3-5 days forecasts of 12 UTC). On the other hand the contradictions between the scores using different verification references indicate that robust conclusions cannot be drawn and all the results should be interpreted with care. It is especially true for the signals beyond 2-3 days, where it is difficult to imagine how a positive (or neutral) influence at the beginning of integration is transformed into a negative one especially considering the small amount of data differences between the two experiments. Based on all these results it is recommended to perform a longer experimentation with the hope for getting more robust conclusions.

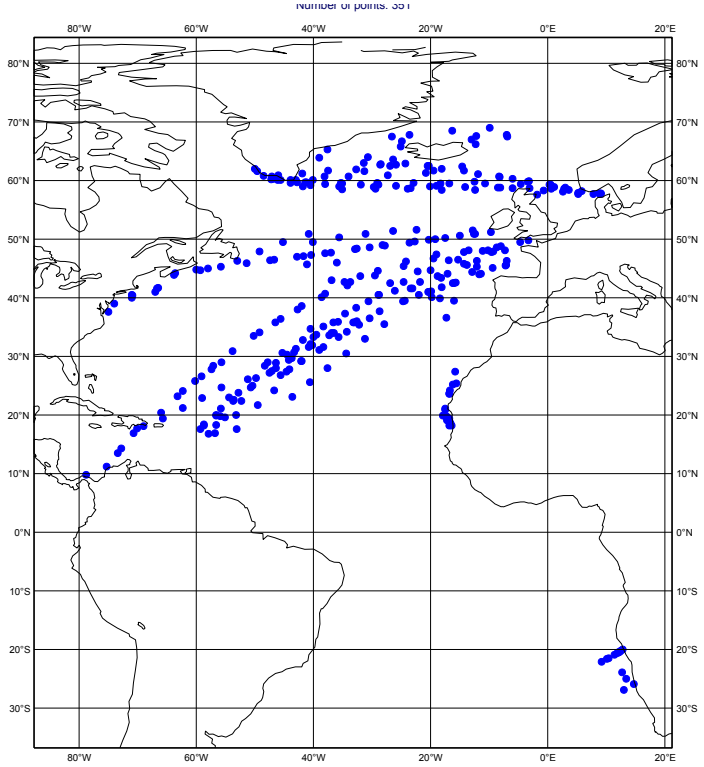


Figure 61: The location of ASAP ships over the Northern Atlantic for the entire 3 months (July-September, 2011) study period.

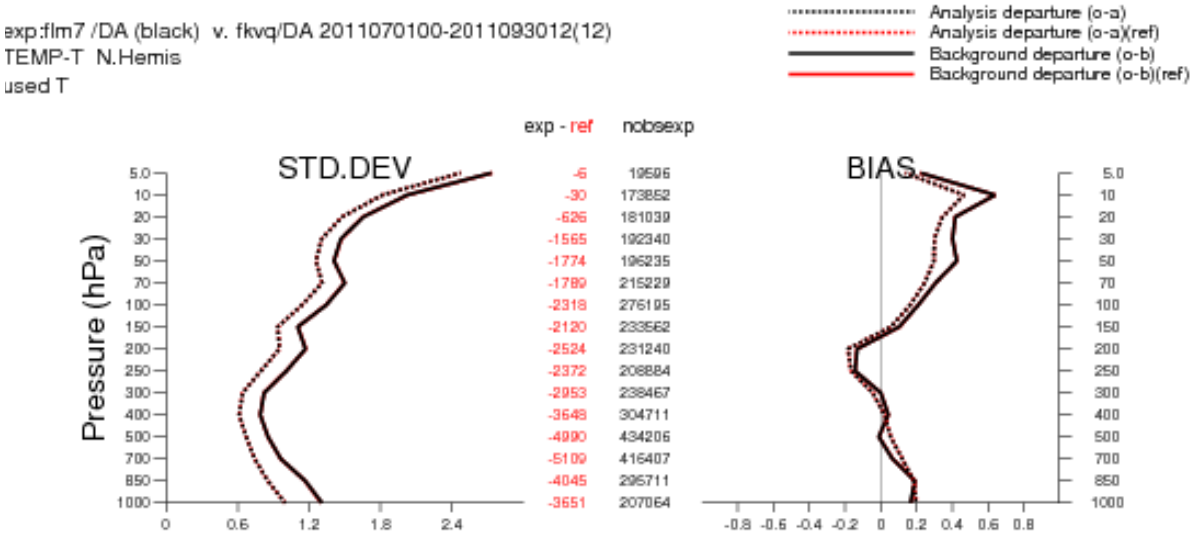


Figure 62: The available radiosonde data in the control and denial experiments (at the middle of the figure, where red numbers indicate the data, which are removed in the denial integrations). The fit of radiosonde observations to the background (continuous line) and analysis (dotted lines) fields in the control (red) and denial (black) experiments, respectively. The left image is standard deviation and the right one is the bias.

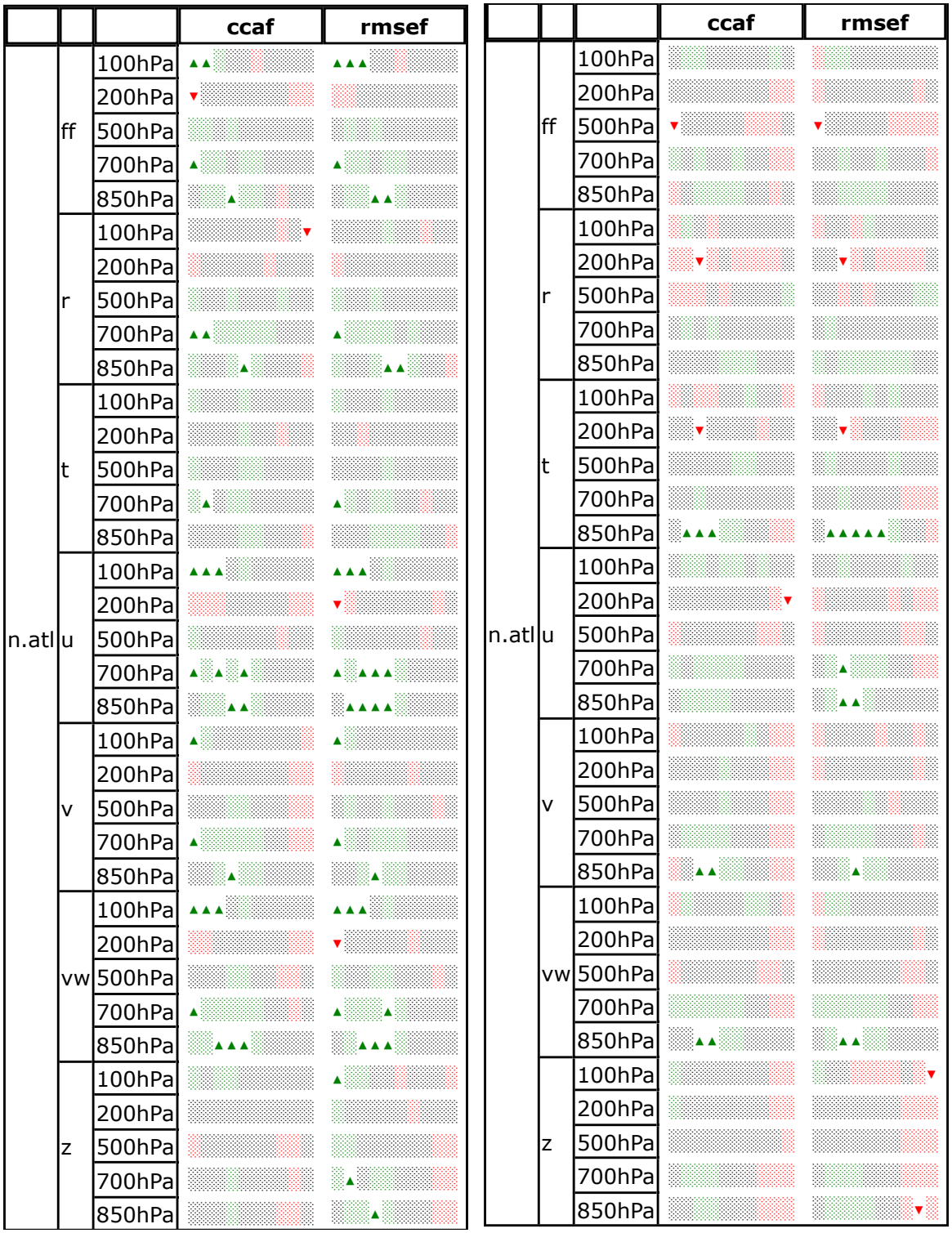


Figure 63: Scorecard of anomaly correlation and RMS error between control and ASAP denial experiments at the Northern Atlantic for different variables, vertical levels and forecast ranges with respect to the IFS operational analysis (left columns) and with respect to the own analysis of the experiments (right columns). Green color represents positive forecast impact by ASAP data. Triangles represent significance of the impact, the bigger the triangle, the stronger the significance. Figure spans the 12h - 120h forecast range. Statistics are derived from 00 and 12 UTC forecasts from the period of 01/07/2011-30/09/2011. The levels, the variables and the score names are indicated on the figure.

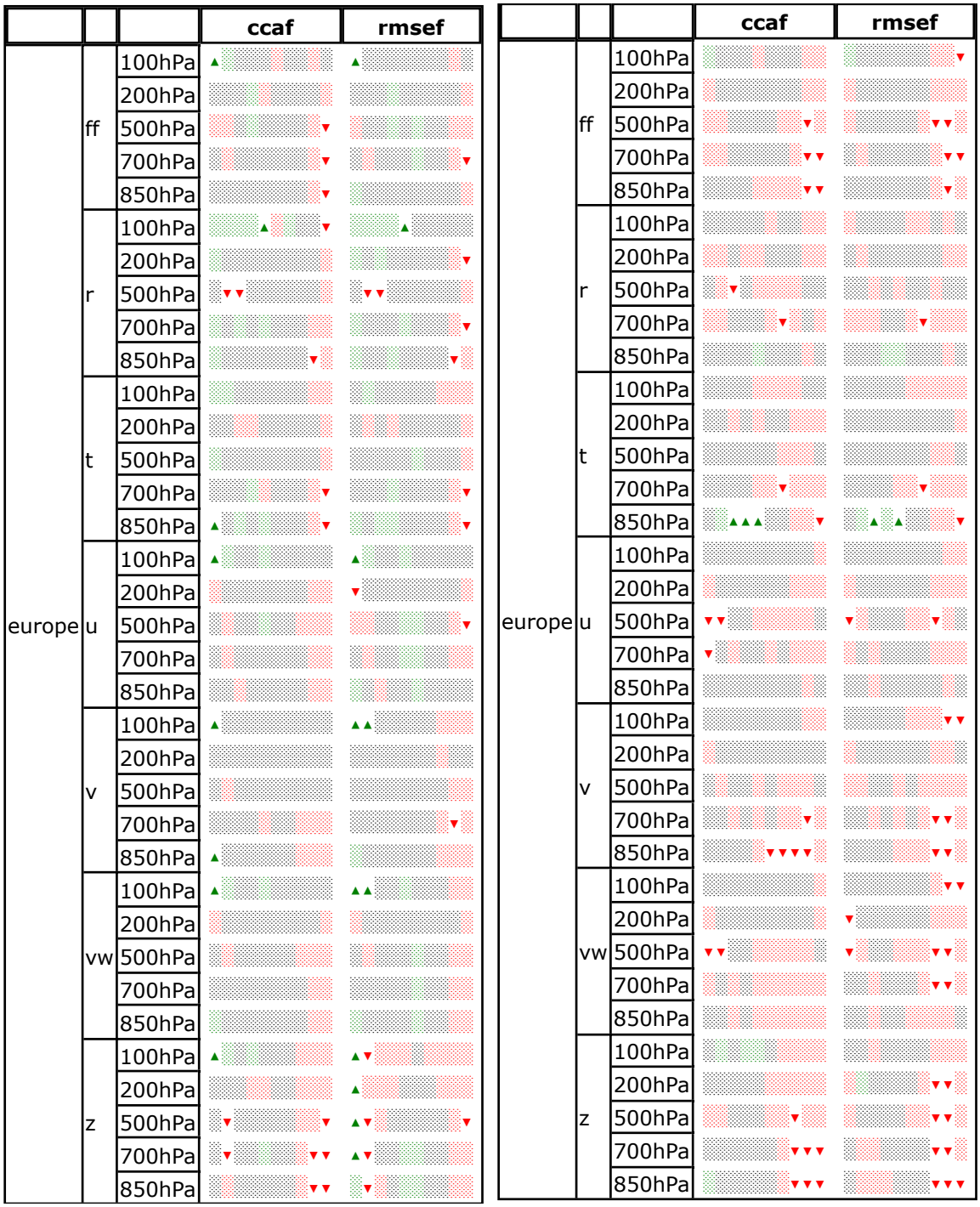


Figure 64: Scorecard of anomaly correlation and RMS error between control and ASAP denial experiments over Europe for different variables, vertical levels and forecast ranges with respect to the IFS operational analysis (left columns) and with respect to the own analysis of the experiments (right columns). Green color represents positive forecast impact by ASAP data. Triangles represent significance of the impact, the bigger the triangle, the stronger the significance. Figure spans the 12h - 120h forecast range. Statistics are derived from 00 and 12 UTC forecasts from the period of 01/07/2011-30/09/2011. The levels, the variables and the score names are indicated on the figure.

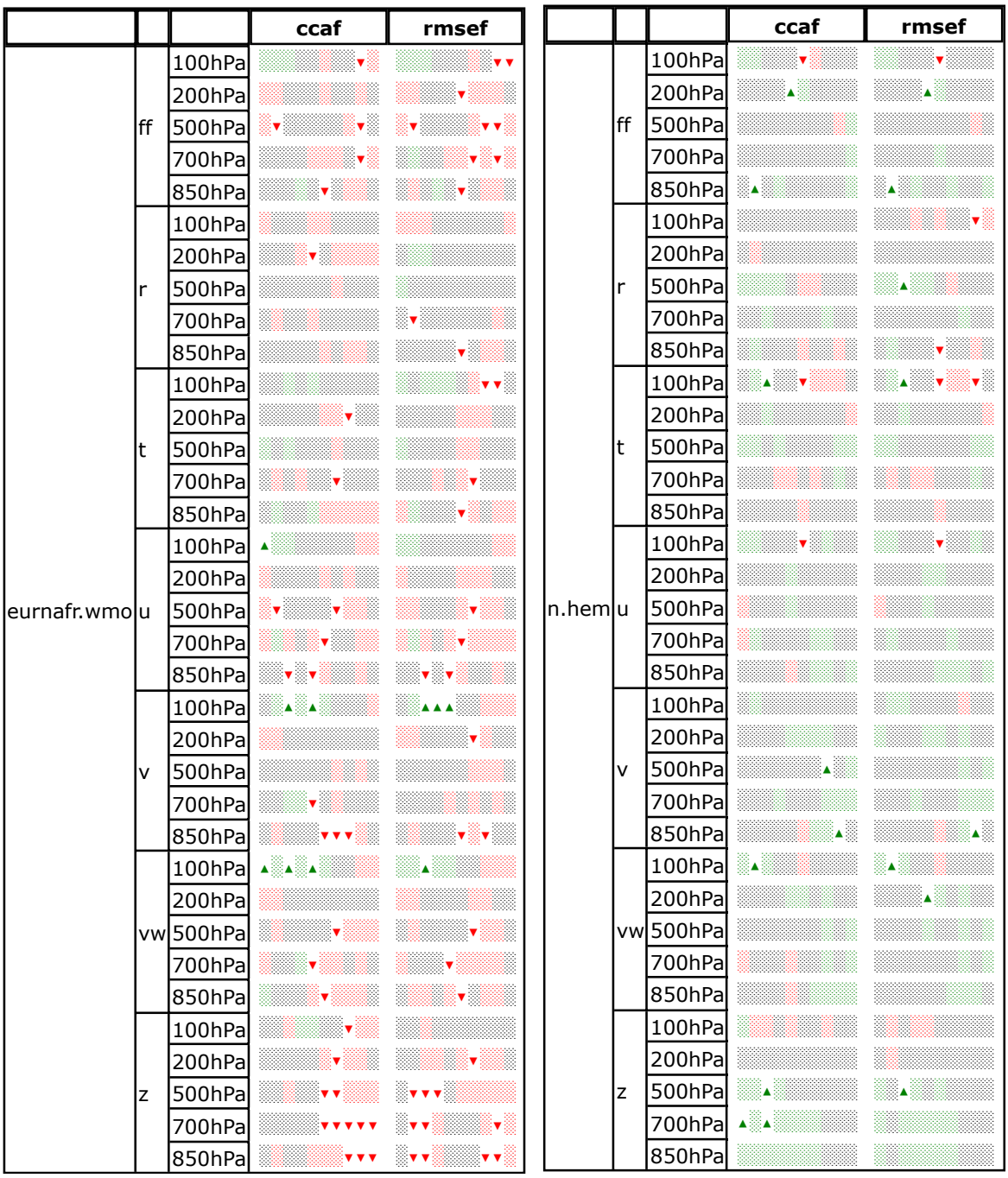


Figure 65: Scorecard of anomaly correlation and RMS error between control and ASAP denial experiments for different domains, variables, vertical levels and forecast ranges with respect to observations. Green color represents positive forecast impact by ASAP data. Triangles represent significance of the impact, the bigger the triangle, the stronger the significance. Figure spans the 12h - 120h forecast range. Statistics are derived from 00 and 12 UTC forecasts from the period of 01/07/2011-30/09/2011. The domains, levels, the variables and the score names are indicated on the figure.

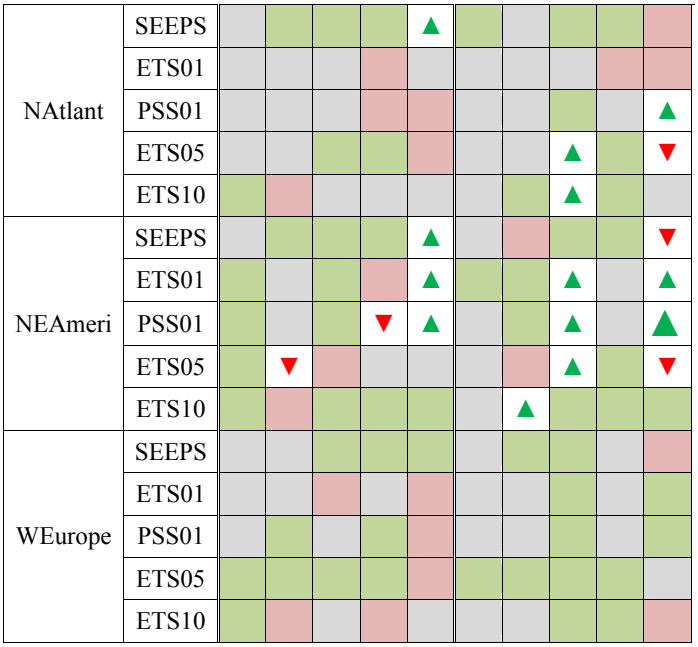


Figure 66: Summary of verification results for ASAP denial experiment vs control. Columns show individual differences for forecast days 1-5 for each model run. Red (green) triangles indicate a positive (negative) score impact of the additional data, difference of 0.01 or more, red (green) shading indicates a positive (negative) score difference of 0.003-0.009, grey shading indicates a score difference of 0.002 or less. Numbers next to ETS and PSS denote thresholds in mm.

6 Acknowledgements

Present study has been funded by EUCOS (Project Reference NC149, EUCOS Reference: EUCOS-ADM-2010-001). The authors wish to thank Stefan Klink, Lars Isaksen and Sean Healy for their comments and many discussions throughout the study. Special thanks to Thomas Haiden for his help in performing precipitation verifications and to Pierre Blouch for providing us with the North-Atlantic marine buoy blacklists. The GPSRO impact studies have been carried out using the RO data reprocessed by UCAR. We wish to thank UCAR for providing us with the dataset.

7 List of figures and tables

List of Figures

List of Tables

8 References

Auligne, T., A.P. McNally, and D.P. Dee, 2007. Adaptive bias correction for satellite data in a numerical weather prediction system. Quart. J. Roy. Meteorol. Soc., 133, 631-642.

Bauer, P. and G. Radnoti, 2009: Study on Observing System Experiments (OSEs) for the evaluation of degraded EPS/Post-EPS instrument scenarios. Report available from ECMWF, Reading, UK, 99pp.

Cardinali, C., 2009a: Forecast sensitivity to observations (FSO) as a diagnostic tool. ECMWF Technical Memorandum, No.599. pp 26. Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.

Cardinali, C., 2009b: Monitoring the observation impact on the short-range forecast. *Q. J. Roy. Meteor. Soc.*, **135**, 239-250.

Dee, D.P., 2005: Bias and data assimilation, *Q. J. Roy. Meteor. Soc.*, **131**, 3323-3343.

Radnoti, G., 2010: Study on Observing System Experiments (OSEs) for the evaluation of degraded European radiosonde and AMDAR scenarios for the EUCOS Operational Programme - Upper-air network redesign. Study Report, available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Radnoti, G., P. Bauer, A. P. McNally, C. Cardinali, S. Healy and P. deRosnay, 2010: ECMWF study on the impact of future developments of the space-based observing system on Numerical Weather Prediction. ECMWF Technical Memorandum, No. 638, 117 pp. Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.

Radnoti, G., 2011: Forecast obstat: A new prepIFS configuration to compare forecast against observations and to store the information on the ODB. ECMWF Technical Report, No. 1108, February, 2011. Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom

Haseler, J., 2004: The early-delivery suite. ECMWF Technical Memorandum, No. 454, 35 pp. Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.

Rodwell, M. J., D. S. Richardson, T. D. Hewson, and T. Haiden, 2010: A new equitable score suitable for verifying precipitation in numerical weather prediction. Q. J. R. Meteorol. Soc., 136, 1344-1363.

Thepaut, J.-N. and G. Kelly, 2007: Relative contributions from various terrestrial observing systems in the ECMWF NWP system. Final ECMWF report of the EUCOS Space Terrestrial Study. Available from EUCOS.

Vasiljevic, D., E. Andersson and A. Garcia-Mendez, 2005: Surface pressure analysis mods in 29r1: bias correction, use of METAR, non-use of PAOBS & reduction of obs errors. ECMWF RD Memorandum, No. 542.

Vasiljevic, D., E. Andersson and A. Agusti-Panareda, 2007: Radiosonde temperature and humidity bias correction for operations, research and reanalysis. ECMWF RD Memorandum, No. 768.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 464pp.